AD_____

Award Number: W81XWH-04-1-0512

TITLE: Promoter and Cofactor Requirements for SERM-ER Activity

PRINCIPAL INVESTIGATOR: Jason S. Carroll, Ph.D.

CONTRACTING ORGANIZATION: Dana-Farber Cancer Institute
Boston MA 02115

REPORT DATE: May 2007

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-05-2007	Final	15 Apr 04 – 14 Apr 07

4. TITLE AND SUBTITLE
Promoter and Cofactor Requirements for SERM-ER Activity

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0512

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jason S. Carroll, Ph.D.

E-Mail: jason.carroll@cancer.org.uk

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Dana-Farber Cancer Institute
Boston MA 02115

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT:
We aimed to understand the mechanisms by which Estrogen Receptor can regulate transcription, both in a positive context after estrogen stimulation and in a negative context after tamoxifen. We originally planned to develop a methodology for specifically isolating chromatin to assess associated proteins, but due to technical limitations, we developed the ChIP-on-chip technique to map ER binding sites, initially on a chromosome-wide scale, but eventually on a genome-wide level. These experiments revealed surprising insight into ER action, namely, that ER rarely binds to promoter sequences, but in a majority of cases binds to regions distant from transcription start sites. We also found an enrichment of motifs that gave us insight into the factors involved in augmenting ER activity. This led to the identification of FoxA1 as a pioneer factor, which we subsequently showed to be an essential component in recruiting ER to the chromatin. The whole-genome analysis revealed another class of cofactors including Oct-1, C/EBPa and AP-1 proteins. We used this information to identify the mechanisms by which ER can repress gene transcription. This included physiologic squelching at the early time point and direct repression via AP-1 elements at the later time points. This study revealed exceptional information about ER function and provides the first complete list of the cis-regulatory elements that may be involved in breast cancer resistance to tamoxifen.

15. SUBJECT TERMS
Estrogen Receptor, transcription, chromatin, SERM

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	41	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

The Estrogen Receptor (ER) status correlates with greater than 60% of breast cancers, where it functions, not only as a marker to grade cancers, but it is the transcription factor that drives cell division. Until 5 years ago, most work attempting to understand ER transcription focused on one or two target genes, including IGF-I, c-Myc and pS2/TFF-1. Reporter assays and gel shifts suggested that the promoter regions of these genes were important for gene transcription and specific motifs or elements were highlighted as essential domains. These included motifs for Sp-1, AP-1 and cAMP factors. However, it is becoming clear that a fragment of DNA behaves differently when in a histone-free plasmid, relative to a natural chromatin context and this has permitted a re-analysis of the conclusions of motifs required for transcription of key target genes. Our understanding of ER biology was revolutionized by the advent of Chromatin Immunoprecipitation (ChIP), which allowed for in vivo identification of ER association with promoter regions. ChIP assays not only clarified the proteins that can bind with ER to promoter regions but showed that these proteins (including ER) can cycle on and off of the chromatin with predictable kinetics.

The major limitation of ChIP assays is that they are restricted to one or two promoter regions that are suspected ER binding sites, since specific primer sequences are required for PCR. We aimed to circumvent this limitation, by combining ER ChIP with microarrays that cover either entire chromosomes (chromosomes 21 and 22) or the entire human genome with tiling properties, which is essentially contiguous 25bp probes end to end along the non-repetitive sequence.

Body

Previously reported work

The ultimate goal of the project was to identify novel proteins that interact with the ER complex during transcription, using in vivo Chromatin Immunoprecipitation (ChIP) assays with novel approaches for identifying proteins. We initially aimed to generate MCF-7 (breast) and ECC1 (endometrial) cancer cells with a single Lox-Luciferase integration cassette embedded within the chromatin, that could be used as an entry point for introduction of promoters of interest. These promoters, included c-Myc, EBAG9, TFF-1 and IGF-1, would be assessed for transcriptional activity (as assessed by luciferase activity) and this transcriptional activity could be monitored when various mutants of the promoter sequences were re-introduced into the same locus of the chromatin. These promoters had previously be cloned into luciferase reporter assays and shown to possess potent transcriptional activity in this histone-free in vitro assay. The secondary goal was to tag the promoters of interest and to subsequently use the tag to precipitate the DNA and assess what proteins are associated with it, in order to identify, in an unbiased manner, the proteins that bind with ER and potentially function as coactivators to augment transcription. We previously reported that we had generated several MCF-7 clonal cell lines and ECC1 clonal cell lines and screened them for the presence of a single integration site. Furthermore, we generated the cloning vectors required for introduction of various promoter regions of interest into the chromatin. We performed these experiments and selected clonal cell lines that contained c-Myc, EBAG9, TFF-1 and IGF-1 promoter regions, to establish individual cell lines that had the different promoter regions in the same chromatin context. However, when we assessed luciferase activity in any of the cell lines, we could not detect any transcription activity under any conditions, including hormone depletion, estrogen addition and growth factor stimulation. This was the case for all the different clonal cell lines and suggested that either the cassette had integrated (in all cases) into a region of the chromatin that was not conducive to transcriptional activity, or alternatively that the 1kb promoter regions could not induce transcription in these chromatin conditions. To identify the mechanisms for this failure of transcriptional activity, we introduced the CMV promoter sequence into the Lox-

integration site in the chromatin and select cells to generate stable clonal cell line that contained the potent CMV promoter in the Lox-Luciferase cassette. When we assayed for luciferase activity using this powerful promoter, we could not detect activity in any MCF-7 clones and only marginally detected activity in one ECC1 clonal cell line. This suggested that in a chromatin context, small DNA sequences with in vitro activity cannot function appropriately. In order to establish if new clonal cell lines could be derived that contained the random Lox-luciferase cassette integrated into a more euchromatic regions that may be more permissive of transcription, we re-transfected in the Lox-Luciferase cassette, selected cells, generated clonal cell lines and assessed them for activity by recombining the CMV promoter into the Lox-luciferase site. None of the newly generated clones possessed any transcriptional activity, negating the ability of this approach to assess the transcriptional activity from specific piece of DNA. Due to this limitation, it was no longer possible to pursue the later aims of identifying essential DNA motifs for transcription and subsequently identifying novel cofactors during ER-mediated transcription. To circumvent this problem we attempted to achieve the same original goal by combining ChIP with microarrays that cover significant regions of unexplored sequence in order to find genuine in vivo ER binding sites that could subsequently be mined to find enriched DNA binding elements and shed light on the unknown cofactors that augment ER transcription.

Development and validation of ER ChIP and amplification of DNA

MCF-7 breast cancer cells are used as a model to understand ER action. We grew MCF-7 cells in complete media and subsequently depleted them of serum by treating for 3 days in Charcoal Dextran Treated (CDT) media. This hormone depleted media results in cell cycle arrest, which was assessed by flow cytometry. Estrogen was added for increasing time periods and the cells were fixed in formaldehyde to maintain protein-protein and protein-DNA interactions, after which chromatin was collected and a specific antibody to ER was used to immunoprecipitate ER, the associated proteins and interacting DNA fragments. The DNA was purified and real time PCR was performed using primers against the promoter of TFF-1, a well-characterized estrogen target gene. The data was normalized to DNA content and further normalized to total genomic DNA (Input) to

assess the enrichment of TFF-1 promoter bound to ER at the different time points of estrogen treatment. A cyclic association of ER was observed, with a maximal recruitment of ER at 45 minutes.

We used DNA bound to ER at the 45 minute time point as a source of chromatin to identify ER binding sites. Due to the low yield of DNA during ChIP (approximately 1 to 2ng), but the large amount of DNA required for microarray analysis (several ug), DNA amplification was required. We utilized a ligation-mediated PCR approach (LM-PCR) that involved a number of steps: 1. Validated DNA was end filled to generate blunt ends, 2. pre-annealed linkers were ligated onto the ends of the DNA fragments in a random manner to generate similar ends on each DNA fragment, 3. limited PCR was performed using a primer against the linker region to amplify the DNA, 4. DNA was purified, quantitated and validation of enrichment was performed using TFF-1 as a positive control. Once the DNA was assessed and shown to be abundant with maintenance of ER binding enrichment on tested sites, we end labeled the DNA using dNTP-biotin and prepared the samples for microarray hybridization.

ChIP-on-chip discovery of ER binding sites and interacting proteins on chromosomes 21 and 22

The microarrays used were generated by Affymetrix and cover the entire non-repetitive DNA sequences of chromosomes 21 and 22 using 25 bp probes every 35 bp across the entire chromosomes (for the methodology refer to attached manuscript Carroll 2005). This results in approximately 1 million probes that cover 35 million bp, including all the genes, introns, and intergenic sequences of chromosomes 21 and 22. These probes are split on a 3 microarray set in order to cover this large region of the genome. As a positive control, TFF-1, the previously validated estrogen target gene is located on chromosome 21. The DNA associated with ER by ChIP was hybridized to the microarrays and data was analyzed by comparing the signal from each Perfect Match (PM) probe and control Mismatch (MM) probes. Once this difference was established, non-parametric Wilcoxin ranked sum analysis was performed using a sliding window of 600bp to identify clusters of positive probes that represent ER binding sites. This analysis involves some simple

parameters, which included the requirement for multiple adjacent probes to be positive and for gaps of a maximum size to limit peak identification. This resulted in 57 ER binding sites on chromosomes 21 and 22 (*1*). As an example, we found ER binding at the promoter of TFF-1, exactly 400bp upstream of the transcription start sites, where a well defined ERE was located (Figure 1). Surprisingly however, we also found an ER binding site 10.5 kb upstream from TFF-1 gene suggesting it may be an enhancer.

To validate some of the newly identified ER binding sites, we designed primers against the chromosomal co-ordinates that were defined as ER binding site peaks and performed standard ER Chip followed by real time PCR of the newly identified sites. All of the sites we tested proved to be genuine in vivo ER binding sites, confirming the power of the ChIP-on-chip approach. We found unique ER binding patterns near several genes of interest, including 10 ER binding sites in the middle of the DSCAM-1 gene, 6 ER binding sites more than 150kb from the transcription start site of the Nuclear Receptor cofactor, NRIP-1, and 3 ER binding sites 15-25 kb upstream of the XBP-1 transcription factor. All of these genes were shown to be estrogen regulated. Furthermore, we performed ChIP using antibodies against RNA Polymerase II and the ER cofactor AIB-1, both of which were shown to be recruited to the ER binding sites in an estrogen dependent manner. To prove that the ER binding sites that were, in some cases, significant distances from the putative gene targets, we applied a Chromosome Conformation Capture (CCC) approach to identify long distance cis-regulatory elements, which proved successful in two of the three assessed cases, including TFF-1 and NRIP-1. This for the first time confirmed that long distance enhancers are used as primary ER binding sites for transcription.

Using the pool of 57 ER binding sites on chromosome 21 and 22, we screened the sequences for DNA binding motifs that were enriched more than expected by chance and found two such elements, namely an Estrogen Responsive Element (ERE) and a Forkhead motif. The finding of EREs validated the technique and proved that we were in fact finding genuine ER binding sites, but the identification of the Forkhead motif suggested a novel role for Forkhead proteins and ER. A search of all the Forkhead

proteins (there are approximately 40 members known, all of which can bind to the same Forkhead motif that was enriched within the ER binding site) in MCF-7 cells using publically available data revealed the high expression of one Forkhead protein, namely FoxA1, which was also shown to correlate with ER status in breast tumors. Furthermore, FoxA1 was shown by others to bind to other Nuclear Receptors including Androgen Receptor (AR) and Glucocorticoid Receptor (GR), all of which suggested that this was the Forkhead protein most likely to bind Forkhead motifs in our system. We performed ChIP of FoxA1 (as well as several other Forkhead proteins as controls) followed by PCR of a number of the newly identified ER binding sites. This resulted in data showing that FoxA1 binds to approximately 50% of all ER binding sites, but interestingly, unlike most proteins co-operating with ER, FoxA1 was on the chromatin before estrogen addition and dissociates from the DNA after estrogen treatment, coincident with ER loading onto the DNA. Since thousands of predicted ER binding sites (in the form of computationally predicted EREs) occurred on chromosome 21 and 22, but only 57 binding sites were observed, the presence of FoxA1 provided the possibility that this Forkhead protein may dictate where ER can bind to the chromatin. To assess this hypothesis, we designed siRNA against FoxA1 and transfected this siRNA into MCF-7 cells, along with siLuciferase as a control. We subsequently assessed FoxA1 protein levels after siRNA and collected RNA after vehicle or estrogen stimulation. When we assessed the estrogen induced mRNA changes in several estrogen target genes on chromosomes 21 and 22, we observed a significant decrease in estrogen induction when FoxA1 was silenced, suggesting that the newly identified ER co-operating factor, FoxA1, is essential for ER activity. In order to assess whether FoxA1 was required for ER to bind to the chromatin, we performed siFoxA1 silencing and then assessed ER recruitment to a number of tested sites by ER Chip. We found that ER could not bind to DNA in the absence of FoxA1, showing a requirement for FoxA1 in defining where and how ER can bind to the chromatin.

ChIP-on-chip discovery of ER binding sites and interacting proteins on the whole human genome

The significant insight gained by mapping ER binding sites on chromosome 21 and 22 provided the impetus to map all ER binding sites across the entire non-repetitive human genome, which constitutes 1.5 billion base pairs tiled at 35 bp resolution (for detailed methodology refer to (*2*)). We performed both ER and RNA PolII ChIP-chip experiments in triplicate across the entire genome and analyzed the data using a novel program that was developed for Affymetrix tiling array data (*3*). This program termed MAT is the most sophisticated approach for converting Affymetrix ChIP-chip data in clear biologically relevant information. Using MAT, we identified approximately 3,600 ER binding sites (using a stringent statistical cutoff) and almost the same number of RNA PolII binding sites across the human genome (*2*). Surprisingly, each gene appears to have a unique binding profile (Figure 2) suggesting that no single paradigm can be used to describe ER binding patterns. Analysis of the ER and RNA Polymerase II sites revealed a significant degree of sequence conservation with the binding sites, suggesting that these discrete regions are conserved in multiple species, highlighting their biological significance during evolution.

To address the major goal of this proposal, we again attempted to identify proteins that would co-operate with ER to mediate transcription, although the current approach used a statistical enrichment of transcription factor binding sites within the newly identified ER binding sites. When we performed this analysis of all 3,665 ER binding sites, we find EREs and Forkhead motifs, as previously identified from chromosome 21 and 22 analyses. However, we also find C/EBP, AP-1 and Oct elements enriched with the ER binding sites, suggesting that the factors that bind to these elements likely contribute, to some degree, to ER transcription. As such, we performed Chip of C/EBPα, Oct-1 and c-Jun (which binds AP-1 motifs) followed by real time PCR of a number of newly discovered ER binding sites. We find C/EBPα, Oct-1 and c-Jun binding to a number of ER binding sites (Figure 3). We designed siRNA to each of these newly implicated factors and showed that by specifically silencing each, we would partially abrogate the estrogen induction of a number of target genes.

To correlate binding events with gene transcription events, we performed expression microarray analysis after an estrogen time course, which included 0, 3, 6 and 12 hours. The 3hr targets are likely to be direct transcriptional targets and the 12 hr targets are likely to be indirect or secondary gene targets. We investigated novel mechanisms of ER gene regulation and found two such possibilities. The genes that were downregulated at the early 3hr timepoint did not have ER binding sites adjacent (within 100kb) of them and we showed experimentally that these genes are down regulated due to physiologic squelching. However, we identified an enrichment of ER binding sites near genes that are downregulated at the later timepoint of 12 hours. These binding sites had an enrichment of AP-1 motifs, whereas the binding sites near the genes that are upregulated at 12 hours have a bias of EREs. We went on to show that NRIP-1 is transcribed at 3hr and then subsequently binds to ER-AP-1 complexes and directly represses gene transcription events. This was shown by silencing NRIP-1 that resulted in an inhibition of the estrogen down regulation of a number of late target genes (*2*). These studies provided novel insight into how ER can turn on genes, but also how ER can down regulate genes and has clear clinical importance, in that it established the first set of co-operating factors (FoxA1, Oct-1, C/EBP etc) and the cis-regulatory elements, which may constitute the elements that cancers mutate to acquire hormone resistance.

Key Research Accomplishments

- Established optimal condition for ChIP-chip experiment

- First ER ChIP-chip experiment, successfully mapping ER binding sites on Chromosomes 21 and 22.

- Defined for the first time, that ER does not bind to promoter regions often, but instead binds to enhancers that are very distant from transcription start sites

- Established Chromosome Conformation Capture method to show that distant enhancer interacts with promoter regions.

- Identified FoxA1 as a pioneer factor for ER binding to the chromatin. The first time such a pioneer factor has been shown to be required for a Nuclear Receptor

- Performed first ER Chip-on-chip on the whole human genome tiling microarrays

- Mapped all ER and RNA PolII binding sites on a genome-wide scale and correlated with gene expression information

- Identified a number of in vivo co-operating factors including Oct-1, AP-1 and C/EBP.

- Identify two different mechanisms of gene repression, namely a early mechanism that utilizes physiologic squelching and a later mechanism that is a direct gene repression

Reportable outcomes

- Poster presented at Keystone 2004
- Seminar presented at Project Program Grant retreat 2004
- Seminar presented at Project Program Grant retreat 2005
- Poster presented at DOD conference 2005
- Manuscript published in Cell 2005
- Development of new analysis tool for Chip-on-chip data (MAT)
- First map of ER binding on entire genome
- Invited seminar, Novartis Institute for Biomedical Research 2005
- Invited seminar, Biomedicum, University of Helsinki, Finland 2005
- Poster award at Harvard breast cancer symposium 2005
- Invited seminar at Harvard breast cancer symposium 2005
- Poster presented at Keystone 2006
- Review article commission in Molecular Endocrinology 2006
- Poster presented at Harvard breast cancer symposium 2006
- Manuscript published in Nature Genetics 2006
- Review article commissioned in Trends in Endocrinology and Metabolism
- First moderator of Harvard Cistrome Meeting 2006
- Invited speaker at Affymetrix Singapore users meeting 2006
- Co-authorships in journals including Molecular Cell, Genes and Development, PNAS, Nature Cell Biology
- Invited Speaker at BES Society for Endocrinology UK 2007
- Faculty position gained at Cancer Research UK/Cambridge Research Institute with a tenure-track position through University of Cambridge, 2007

- Invited speaker at Affymetrix ChIP-chip symposium 2007
- Invited speaker at Imperial College London 2007
- Review article commission for Nature Reviews Cancer 2007
- Manuscript in process 2007

Conclusions

The work conducted under the DOD Breast cancer fellowship has led to a significant advance in our understanding of Estrogen Receptor (ER) action and for the first time has illuminated clinically relevant elements of the pathway. Using the powerful ChIP-chip technique, we defined new paradigms of ER action, namely that ER rarely binds to promoters, but instead binds to distant enhancers. The data also revealed that a number of co-operating factors are involved in ER activity, including FoxA1 which is required for ER to bind to the chromatin and Oct-1, AP-1 and C/EBP that can assist in transcriptional activity. The whole human genome map of ER binding and activity revealed novel insight into methods by which estrogen can down regulate genes, an area that had not be adequately addressed before. The body of data generated by the DOD Fellowship is an excellent resource that can be mined for many years by people interested in estrogen target genes. It also provides the first map of the cis-regulatory elements that likely allow ER to function in breast cancers and which may constitute the sites of perturbation in hormone independence and tamoxifen resistance. However, the major conclusion is that the ultimate goal set out at the beginning of the fellowship application, namely, the identification of ER associated cofactors and proteins, was exceptionally successful. We identified a critical role for Oct-1 and C/EBPα, identified AP-1 associated factors as key components in gene repression by ER, but most importantly, we identified FoxA1 as a Pioneer factor that is essential for ER to bind to chromatin and induce gene transcription.

References

1. J. S. Carroll *et al.*, *Cell* **122**, 33 (2005).
2. J. S. Carroll *et al.*, *Nat Genet* **38**, 1289 (Nov, 2006).
3. W. E. Johnson *et al.*, *Proc Natl Acad Sci U S A* **103**, 12457 (Aug 15, 2006).

Bibliography

Papers published

Carroll, J.S, Liu, X.S, Brodsky, A.S, Meyer, C.A, Li, W, Szary, A.J, Eeckhoute, J, Shao, W, Hestermann, E.V, Geistlinger, T.R, Fox, E.A, Silver, P.A and Brown, M. Chromosome-wide mapping of Estrogen Receptor binding reveals long-range regulation requiring the Forkhead protein FoxA1. Cell. 2005, 122:33-43

Wang, Q, **Carroll, J.S** and Brown, M. Novel special and temporal recruitment of Androgen Receptor and its coactivators leads to chromosomal looping and polymerase tracking. Molecular Cell, 2005, 19:631-642

Carroll, D.K, **Carroll, J.S**, Sofer, A, Cheng, F, Brown, M, Mills, A.A, Brugge, J.S and Ellisen, L.W. p63 regulates an adhesion program and cell survival in epithelial cells. Nature Cell Biology, 2006, 8:551-61

Carroll, J.S and Brown, M. Estrogen Receptor Target Gene: an Evolving Concept. Molecular Endocrinology, 2006, 20:1707-1714

Johnson, W. E, Li, W, Meyer, C. A, Gottardo, R, **Carroll, J. S**, Brown, M and Liu, X. S. MAT: Model-based analysis of tiling-arrays for ChIP-chip, Proceedings of the National Academy of Sciences, 2006, 103:12457-12462

Eeckhoute, J, **Carroll, J.S,** Geistlinger, T.R, Torres-Arzayus, M.I and Brown, M. Combinatorial transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer cells, Genes and Development, 2006, 20:2513-2526

Carroll, J.S, Meyer, C.A, Song, J, Li, W, Geistlinger, T.R, Eeckhoute, J, Brodsky, A.S, Keeton, E.K, Fertuck, K.C, Hall, G.F, Wang, Q, Bekiranov, S, Sementchenko, V, Fox, E.A, Silver, P.A, Gingeras, T.R, Liu, X.S and Brown, M. Defining the Estrogen Receptor Cistrome in Breast Cancer, Nature Genetics, 2006, 38:1289-97

Wang, Q, Li, W, Liu, X.S, **Carroll, J.S**, Janne, O.A, Chinnaiyan, A.M, Pienta, K.J and Brown, M. A transcriptional regulatory network governs prostate cancer growth and survival, Under Review

Schmelzle, T, Mailleux, A.A, Overholtzer, M, **Carroll, J.S**, Solimini, N.L, Lightcap, E.S, Veiby, O.P and Brugge, J.S. Functional role and oncogene-regulated expression of the BH3-only proapoptotic factor Bmf in mammary epithelial anoikis. 2007, PNAS. In Press

Meeting Abstracts

Keystone, Nuclear Receptors 2004

Boston Project Program Grant retreat 2004

Harvard Breast Cancer Symposium 2005

Boston Project Program Grant retreat 2005

DOD meeting 2005

Harvard Breast Cancer Symposium 2006

Keystone, Nuclear Receptors 2006

Personnel covered

Jason Carroll Postdoctoral Fellow

Appendices

Figure 1

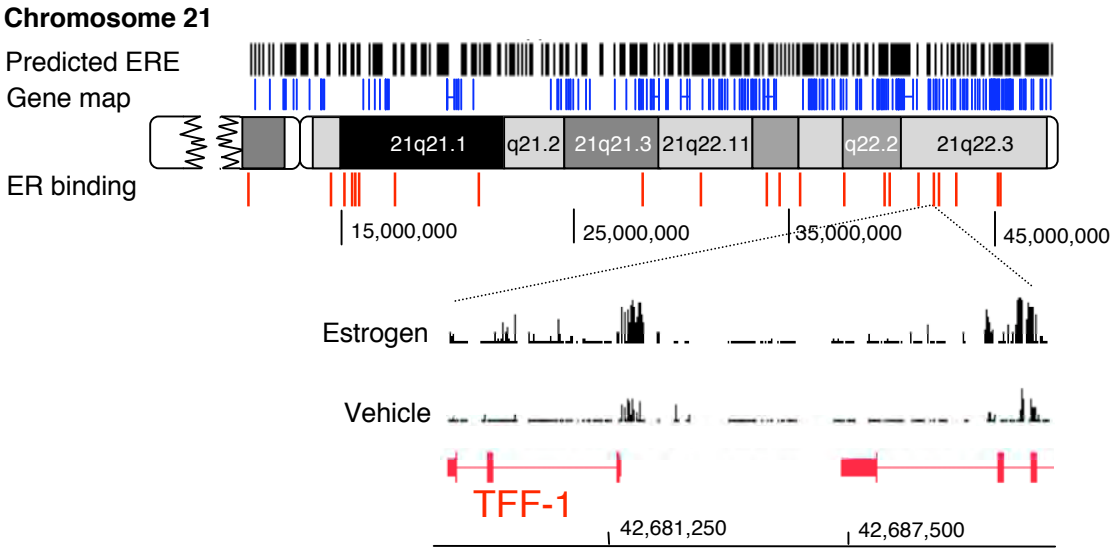
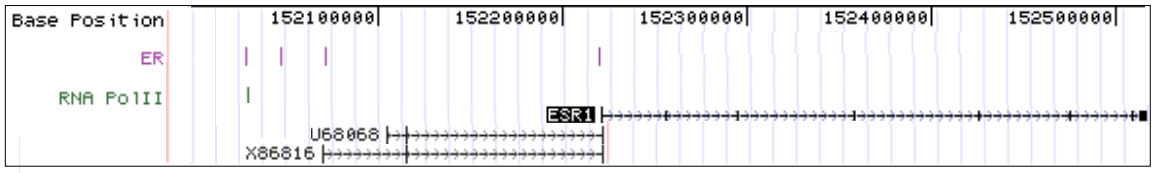


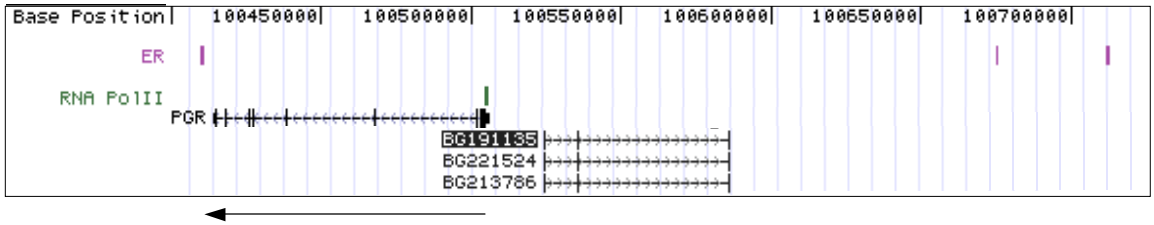
Figure. 1
Map of ER binding sites on chromosomes 21 after estrogen stimulation. Genes locations are shown in blue bars. Gene locations are based on the April 2003 genome freeze in the UCSC browser using Genbank RefSeq positions. Predicted EREs are shown as black bars and ER-binding sites are shown as red bars. An expanded view of the TFF-1 gene region is shown as signal difference between ER ChIP and Input DNA for both the estrogen and vehicle treated cells. The TFF-1 gene is shown in its genuine 3'-5' orientation.

Figure 2

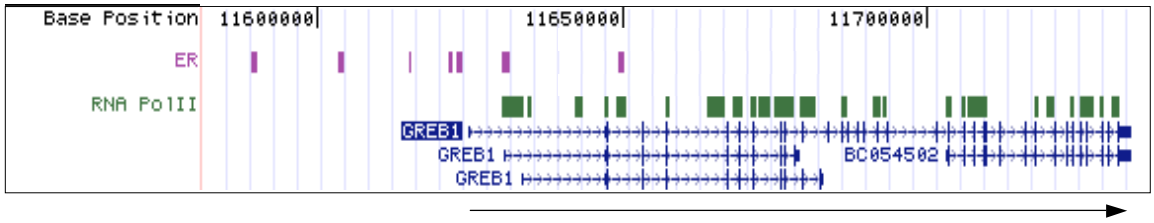
Estrogen receptor (*ESR1*)



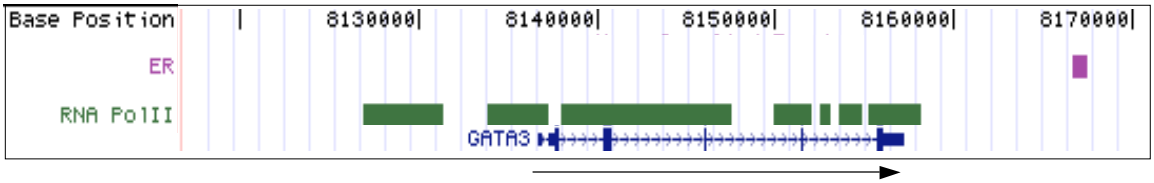
Progesterone receptor (*PGR*)



GREB1



GATA3



MYC

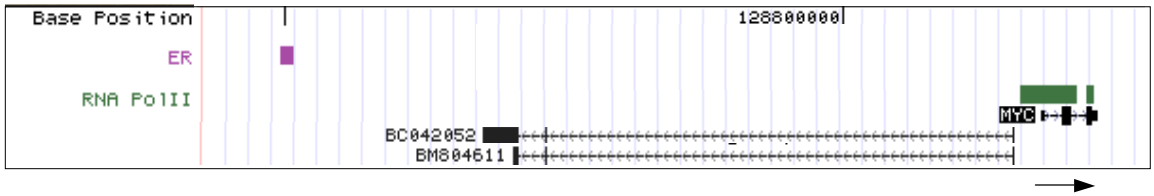
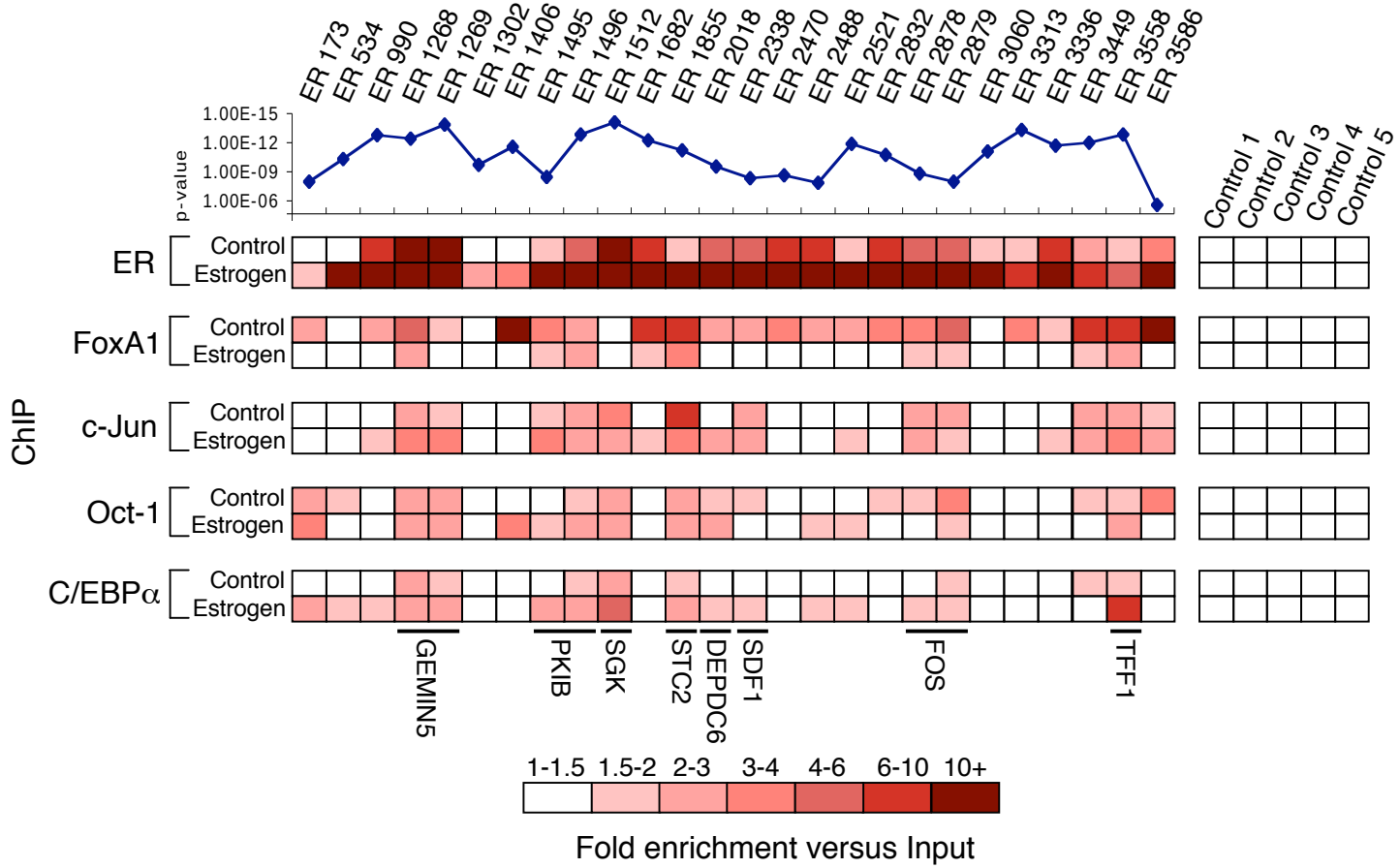


Figure 2.
ER and RNA PolII binding relative to specific genes targets. The purple blocks represent ER binding sites and green blocks represent RNA PolII sites. *Estrogen Receptor*, *GREB-1*, *c-Myc* and *GATA3* are shown in their genuine 5'-3' orientation and *Progesterone Receptor* is shown in its genuine 3'-5' orientation. The black arrows indicate the direction of the gene. Included are predicted transcripts that exist between the ER binding sites and the target genes.

Figure 3



Directed ChIP of transcription factors that bind to enriched motifs was performed on 26 ER binding sites and 5 control regions. The binding sites were chosen to cover a range of enrichment values, but also included sites near a select number of estrogen-regulated genes. The relative p-value for each of the binding sites assessed is provided. ER binding sites adjacent to estrogen-regulated genes are shown by the gene name. The real time PCR data is shown as fold enrichment relative to input DNA and is the average of independent replicates.

Cell, Vol. 122, 33–43, July 12, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.05.008

Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1

Jason S. Carroll,[1] X. Shirley Liu,[2,4] Alexander S. Brodsky,[3,5] Wei Li,[2,4] Clifford A. Meyer,[2,4] Anna J. Szary,[1] Jerome Eeckhoute,[1] Wenlin Shao,[1] Eli V. Hestermann,[6] Timothy R. Geistlinger,[1] Edward A. Fox,[3] Pamela A. Silver,[3,5] and Myles Brown[1,*]

[1]Department of Medical Oncology
[2]Department of Biostatistics and Computational Biology
[3]Department of Cancer Biology
Dana-Farber Cancer Institute
Harvard Medical School
44 Binney Street
Boston, Massachusetts 02115
[4]Harvard School of Public Health
Harvard Medical School
677 Huntington Avenue
Boston, Massachusetts 02115
[5]Department of Systems Biology
Harvard Medical School
200 Longwood Avenue
Boston, Massachusetts 02115
[6]Department of Biology
Furman University
3300 Poinsett Highway
Greenville, South Carolina 29613

Summary

Estrogen plays an essential physiologic role in reproduction and a pathologic one in breast cancer. The completion of the human genome has allowed the identification of the expressed regions of protein-coding genes; however, little is known concerning the organization of their *cis*-regulatory elements. We have mapped the association of the estrogen receptor (ER) with the complete nonrepetitive sequence of human chromosomes 21 and 22 by combining chromatin immunoprecipitation (ChIP) with tiled microarrays. ER binds selectively to a limited number of sites, the majority of which are distant from the transcription start sites of regulated genes. The unbiased sequence interrogation of the genuine chromatin binding sites suggests that direct ER binding requires the presence of Forkhead factor binding in close proximity. Furthermore, knockdown of *FoxA1* expression blocks the association of ER with chromatin and estrogen-induced gene expression demonstrating the necessity of FoxA1 in mediating an estrogen response in breast cancer cells.

Introduction

Estrogen is an essential regulator of female development and reproductive function and has been implicated as a causal factor in breast and endometrial cancers. Estrogen-regulated gene expression is mediated by the action of two members of the nuclear receptor family, ERα and ERβ, with ERα being dominant in both breast epithelial cells and in breast cancer. Significant progress has been made over the past decade in defining the complex interactions between chromatin and an array of factors involved in ER-mediated gene expression (Halachmi et al., 1994; Metivier et al., 2003; Shang and Brown, 2002; Shang et al., 2000), including the cyclic association of ER, p160 coactivators (such as AIB-1), histone acetyl transferases (HAT), and chromatin modifying molecules, such as p300/CBP and p/CAF, with target promoters in an ordered temporal fashion (Metivier et al., 2003; Shang et al., 2000).

In addition, a number of recent strategies including gene expression profiling on microarrays have identified potential ER target genes in human breast cancer cells and only a few *cis*-elements targeted directly by ER have been identified to date. For example, estrogen responsive elements (ERE) have been identified within the 1 kb 5′-proximal region of the estrogen-regulated genes *TFF-1 (pS2)*, *EBAG9*, and *Cathepsin D* (Augereau et al., 1994; Berry et al., 1989; Ikeda et al., 2000), and the proximal promoters of target genes that lack EREs, including *c-Myc* and *IGF-I*, contain AP-1 and Sp-1 sites that appear essential for transcription in in vitro reporter assays (Dubik and Shiu, 1992; Umayahara et al., 1994). Few, if any regulatory elements at significant distances from the mRNA start sites of target genes have been shown to be directly targeted by ER, and computation approaches to identify novel ER binding domains have focused primarily on gene proximal regions (Bajic and Seah, 2003; Bourdeau et al., 2004). However, more progress has been made in studies of *β-globin* gene regulation which has contributed to our understanding of general mechanisms of transcriptional regulation and has shown that locus control regions (LCR) up to 25 kb from the gene are capable of enhancing gene transcription (recently reviewed in Bulger et al. [2002]). In this study, we have undertaken an unbiased approach to identify all regulatory regions that may play a role in ER-mediated transcription by combining chromatin immunoprecipitation (ChIP) analyses of in vivo ER-chromatin complexes with Affymetrix tiled oligonucleotide microarrays that cover the entire nonrepetitive sequences of chromosomes 21 and 22, including, importantly, all the intergenic regions. Most previous ChIP-microarray studies have focused primarily on promoter regions (Odom et al., 2004) or CpG islands, which represent promoter-rich sequences (Weinmann et al., 2002). The tiled arrays used here are composed of 25 bp probes located at 35 nucleotide resolution (Cawley et al., 2004; Kapranov et al., 2002) and permit the opportunity to interrogate previously unexplored regions of chromosomal DNA. The 780 characterized or predicted genes on chromosomes 21 and 22 represent about 2% of the total number of genes (Kapranov et al., 2002) and thus provide a representative model for

*Correspondence: myles_brown@dfci.harvard.edu

the unbiased identification of ER-mediated gene regulation paradigms.

Here we find a discrete number of ER binding sites across chromosomes 21 and 22, almost all of which are in nonpromoter proximal regions. We explored underlying biological patterns within the list of genuine chromatin-interacting domains and identified common motifs highly enriched in these regions. Using this information, we prove that the distal ER binding sites are discrete chromatin regions involved in transcriptional regulation and that a Forkhead protein, at these sites, is required for activity by ER.

Results

ER Occupies a Limited Number of Binding Sites on Chromosomes 21 and 22

Estrogen-dependent MCF-7 breast cancer cells were deprived of hormones and stimulated with estrogen or vehicle for 45 min, a time we have previously shown to have maximal recruitment of ER to the promoters of several known gene targets, including *Cathepsin D* and *TFF-1* (Shang et al., 2000). Following ChIP, ER-associated DNA was amplified using nonbiased conditions, labeled, and hybridized to the tiled microarrays. Relative confidence prediction scores were generated by quantile normalization across each probe followed by an analysis using a two-state Hidden Markov model (Rabiner, 1989). These scores included both probe intensity and width of probe cluster. Triplicate experiments eliminated stochastic false positives, after which peaks that reproducibly appeared at least twice in the three replicates were included. Real-time PCR primers were designed against numerous peaks in the list, and directed ER ChIP was conducted to identify the boundary between the true ER binding peaks (>1.5-fold enrichment over input) and the false positives (data not shown) and generate the final list of 57 estrogen-stimulated ER binding sites within 32 discrete clusters (Figures 1A and 1B and see the Supplemental Raw Data in the Supplemental Data available with this article online).

As one example of the validity of this method, the localization of ER to the proximal promoter 400 bp region of the estrogen-regulated gene, *TFF-1*, was observed. A functional ERE had been previously mapped to the region 393 to 405 bp upstream from the transcription start site of *TFF-1* (Berry et al., 1989). Furthermore, a region 10.5 kb upstream of the *TFF-1* transcription initiation site (Figure 1A) was also found to be bound by ER. Interestingly, an estrogen-inducible DNase I hypersensitive site has been previously mapped 10.5 kb upstream from the *TFF-1* start site (Giamarchi et al., 1999), though the region had not been further characterized. Our data now define these regions as authentic ER binding sites.

Within the small list of 57 ER binding sites, we observed 32 ER binding clusters, some of which were proximal to genes previously implicated as estrogen targets, including the transcription factor *XBP-1, DSCAM-1*, and the nuclear receptor coregulator *NRIP-1* (Cavailles et al., 1995; Pedram et al., 2002; Wang et al., 2004). Binding sites were also observed within 200 kb from genes not previously implicated as estrogen targets, including *SOD-1*, a superoxide dismutase gene involved in scavenging oxygen-free radicals (Beckman et al., 1993; Singh et al., 1998) and implicated in tamoxifen-resistant progression in MCF-7 xenografts (Schiff et al., 2000). None of these genes recruited ER to a proximal 5′ promoter region, but possessed divergent patterns of association. The *XBP-1* gene, recruited ER to three distinct and discrete regions 13.2 kb to 22.9 kb upstream of the transcription start site (Figure 1B). *DSCAM-1* contained a clustering of ten intronic ER binding sites, more than 0.5 Mb from the transcription initiation site. *NRIP-1* contained six ER binding sites in a region of chromosome 21 well known for its scarcity of genes (Katsanis et al., 1998). 5′ RACE was performed on *NRIP-1* to determine the exact location of the transcription start site and the distance between the ER binding sites and the genuine transcriptional start site. Sequencing of the 5′ terminus of the *NRIP-1* transcript after estrogen stimulation revealed the presence of two previously missed exons for *NRIP-1*, 74.96 kb and 97.39 kb from the previously annotated gene start site (data not shown). Therefore, the ER binding domains exist 107 to 144 kb from the genuine transcription start site of *NRIP-1*. The locations of all binding sites in relation to genes can be found in Table S1.

The ER binding sites adjacent to *TFF-1*, *XBP-1*, *SOD-1*, *NRIP-1*, and *DSCAM-1* were validated by ER ChIP and standard PCR (Figures 2A–2E). Also, quantitative PCR was performed on each of these sites after ER ChIP (Figure 2F), confirming these putative in vivo binding sites as genuine ER binding sites. To test whether these discrete ER recruitment regions were unique to estrogen action in MCF-7 cells, we performed ER ChIP and directed real-time PCR against the same sites in T47-D breast cancer cells. These data confirmed that the majority of the sites identified in MCF-7 cells were also regions of estrogen-dependent ER binding in a second ER-positive breast cancer cell line (data not shown), highlighting the conservation of specific ER-chromatin association sites.

A Significant Number of ER Binding Sites Reside Adjacent to Estrogen Gene Targets

Estrogen-mediated transcript changes were identified by converting RNA from vehicle or estrogen-stimulated MCF-7 cells into double-stranded cDNA and hybridizing to the chromosome 21 and 22 tiled microarrays. Thirty-five genes (4.4% of all genes) appeared to be transcribed, after which real-time primers were made against all these transcripts and quantitative RT-PCR showed that 12 transcripts on chromosomes 21 and 22 were estrogen induced (Table 1). Eleven of these twelve genes had ER binding clusters within 200 kb. The only estrogen-regulated gene that did not have an adjacent ER binding cluster was *ATP5J*. *TFF-1*, *XBP-1*, and *NRIP-1* were in the small list of 1.5% of genes upregulated following estrogen stimulation (Supplemental Raw Data). *DSCAM-1* and *SOD-1* were not upregulated by estrogen stimulation at the 3 hr time point assessed but were transcribed after 6 hr of estrogen stimulation, as determined by RT-PCR (Figure S2). This delay between ER association and transcription of *DSCAM-1* and

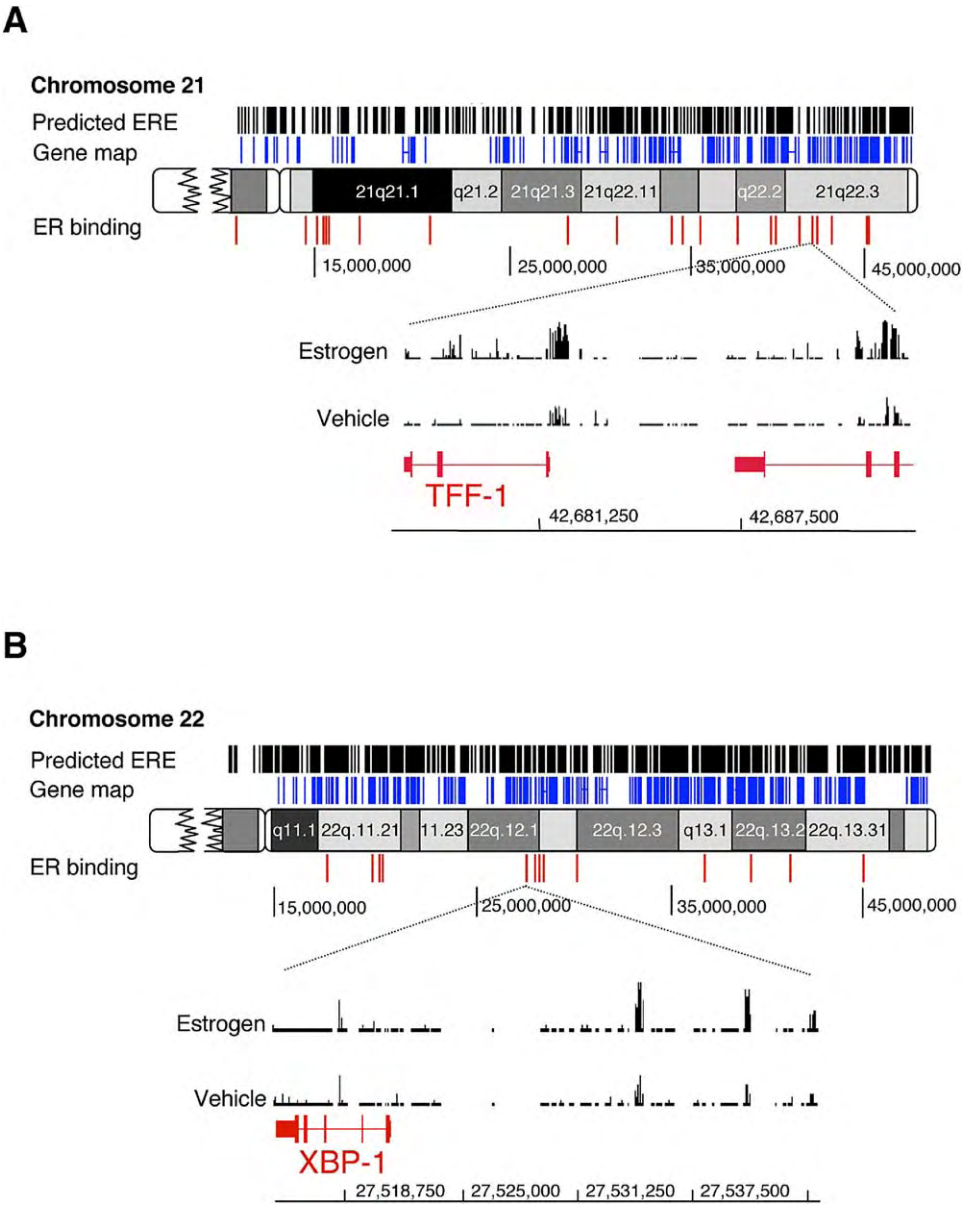


Figure 1. Map of ER Binding Sites on Chromosomes 21 and 22 after Estrogen Stimulation

The visual representation of ER binding sites on chromosomes 21 (A) and 22 (B) are shown. Gene locations are shown in blue bars. Gene locations are based on the April 2003 genome freeze in the UCSC browser using Genbank RefSeq positions. Predicted EREs are shown as black bars and ER binding sites are shown as red bars.

(A) An expanded view of the *TFF-1* gene region is shown as signal difference between ER ChIP and Input DNA for both the estrogen- and vehicle-treated cells. The *TFF-1* gene is shown in its genuine 3′-5′ orientation. The gene adjacent to *TFF-1* is not an estrogen target.

(B) Expanded view of the *XBP-1* gene region on chromosome 22. The *XBP-1* gene is shown in its genuine 3′-5′ orientation.

SOD-1 may be a consequence of a requirement for subsequent modification of the receptor complex or the requirement for the production of other factors involved in ER action but not necessarily part of an ER complex. Regardless of the mechanism for the transcriptional delay, it now appears that early and at least some delayed estrogen-regulated genes recruit the receptor with the same kinetics. This implies that events subsequent to ER binding are responsible for timing the initiation of transcription of these delayed targets.

Distal ER Binding Domains Function as Transcriptional Enhancers

The significant sequence distance between many of the ER binding sites and the putative target gene complicates their functional validation. However, we explored the possibility that these ER binding sites may recruit components indicative of transcriptional activation. RNA PolII ChIP followed by real-time PCR was performed on a subset of the putative regulatory regions adjacent to *TFF-1*, *XBP-1*, *DSCAM-1*, *NRIP-1*, and

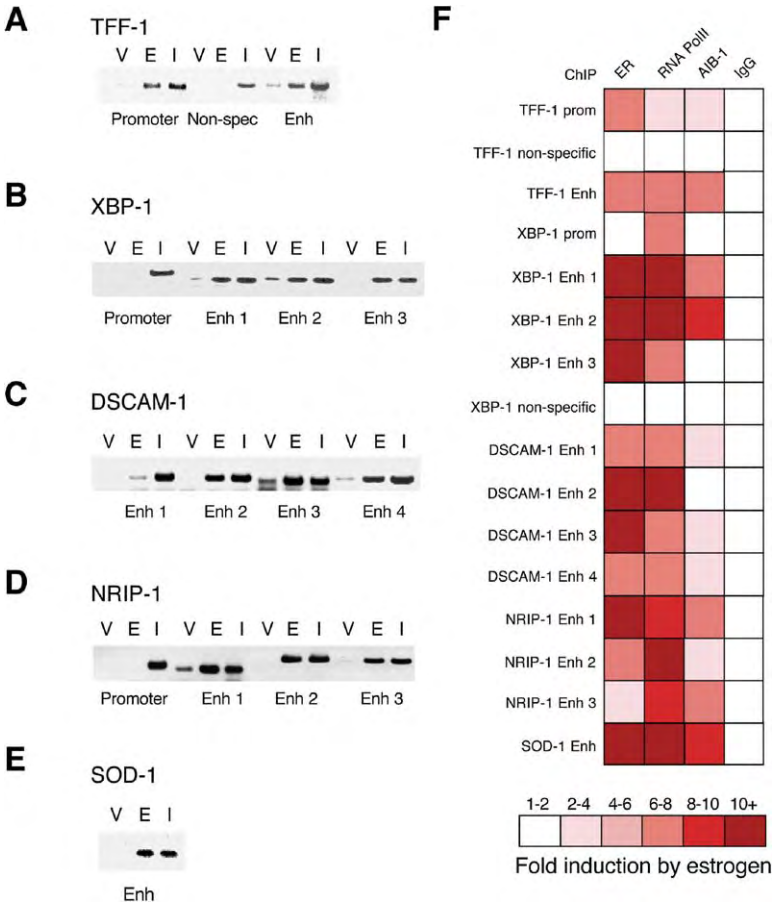


Figure 2. Validation of the In Vivo Binding of the Transcription Complex to Regulatory Regions

ChIP of ER and standard PCR of sites adjacent to *TFF-1* (A), *XBP-1* (B), *DSCAM-1* (C), *NRIP-1* (D), and *SOD-1* (E). *TFF-1* nonspecific and *XBP-1* promoter primers were included as negative controls. The lanes are vehicle (V), estrogen (E), and Input (I).

(F) ChIP of ER, RNA PolII, AIB-1, or IgG control and real-time PCR of binding regions. The data are estrogen-mediated fold enrichment compared to vehicle (ethanol) control. The color intensity reflects the fold change as described in the legend. *TFF-1* nonspecific and *XBP-1* nonspecific primers were included as negative controls. The data are the average of three replicates ± SD.

SOD-1 genes. Interestingly, RNA PolII association was seen with all of these sites in an estrogen-dependent manner (Figure 2F). Furthermore, ChIP of AIB-1, an oncogenic ER coactivator (Kuang et al., 2004; Torres-Arzayus et al., 2004), confirmed that AIB-1 is also present on all of these "regulatory" sites following estrogen exposure (Figure 2F). As negative controls, primers were designed against the intergenic region between the *TFF-1* promoter and enhancer and against a region 7 kb from *XBP-1* enhancer 3. Neither ER nor any of the other factors were found associated with these control regions. In addition, we examined the promoter of *XBP-1*. Although ER protein association was not observed at the *XBP-1* promoter, RNA PolII was found enriched at this site supporting the hypothesis that *XBP-1* is transcriptionally activated by ER.

To explore the possibility that the distal enhancer regions not only function as sites of protein recruitment but physically play a role during transcription of the adjacent gene, we performed a chromosome capture assay (Dekker et al., 2002) to assess whether promoter and enhancer sequences were components of the same chromatin regions. Hormone-depleted MCF-7 cells were stimulated with vehicle or estrogen, and the fixed chromatin was digested with a specific restriction enzyme (BtgI), followed by ER ChIP and ligation. After ligation, the ligated chromatin mix was washed and the crosslinking was reversed. One primer in the *TFF-1* promoter and one primer in the *TFF-1* enhancer were used to PCR potentially ligated fragments of DNA (Horike et al., 2005). As seen in Figure 3A, *TFF-1* promoter and enhancer DNA was ligated together only in the presence of estrogen, confirming that estrogen-mediated transcription of *TFF-1* involves direct physical interaction between the enhancer and promoter. No interaction was seen in the no-digestion control or no-ligation control. We performed the same experiment using the BsmI restriction enzyme that cuts the genuine *NRIP-1* promoter (as determined by 5′ RACE) and enhancer 3 region. Remarkably, after ligation, we were able to PCR a 1 kb fragment that corresponded to the ligated promoter-enhancer regions using one promoter-specific and one enhancer-specific primer (Figure 3B). This estrogen-dependent interaction of the distal (144 kb) ER binding site with the promoter of the *NRIP-1* gene confirms the authenticity of these distal sites as transcriptional regulatory domains.

The finding that RNA PolII is recruited to the majority of ER binding sites, even those removed from known transcription sites, led us to investigate the possibility that these binding sites can function as genuine enhancers. To this end, we cloned 23 ER sites (40% of all ER binding sites) into a pGL-3 luciferase vector containing an SV40 promoter and transfected these vectors into hormone-depleted MCF-7 cells which where subsequently treated with estrogen or vehicle control. pGL-3 empty vector was used as a negative control, and transfections were normalized with pRL null. Al-

Table 1. List of ER Binding Site Clusters and Relative Locations to Putative Gene Targets

Cluster Number	Number of Binding Sites	Start	Stop	Closest Regulated Gene
1	1	21: 10048850	10049271	
2	1	14600251	14600737	
3	1	15171656	15172273	
4	6	15467150	15738864	NRIP-1
5	1	17422343	17422868	
6	1	21532885	21533421	
7	1	29151881	29152882	
8	1	31821967	31822715	SOD-1
9	2	35021165	35027898	
10	1	35510057	35510719	
11	2	36480740	36487032	
12	1	38635468	38636783	
13	10	40363341	40675801	DSCAM-1
14	1	41911683	41912284	
15	1	42005946	42006169	PRDM15
16	2	42680784	42691725	TFF-1
17	1	42830736	42831350	
18	1	43564518	43565261	NDUFV3
19	2	45606461	45663897	
20	1	45790004	45790654	Col18A1
21	2	22: 17159455	17194014	
22	1	19566341	19566809	
23	3	19822950	19945255	
24	3	27534171	27543908	XBP-1
25	1	28106122	28107112	AP1B1
26	1	28237489	28238464	
27	1	28519139	28520023	
28	2	30300284	30307434	PISD
29	2	37030766	37033295	
30	1	39371665	39372232	
31	1	41361325	41361720	Predicted
32	1	45100090	45100552	

The 32 transcriptional clusters are shown, with the start and stop locations of the ER binding sites.

most 75% of the ER binding domains contained estrogen-induced enhancer characteristics in an in vitro transcription model (Figure 3C), supporting the hypothesis that the distal binding sites play transcriptional regulatory roles.

ER Binding Sites Are Conserved Across Species

To identify if the ER binding sites are conserved between human and mouse genomes, we assessed the identity in sequence in a window of 6 kb from the center of all 57 ER binding sites. This conservation was mapped within a 500 bp window at a single nucleotide resolution and confirms a strong conservation at the center of the ER binding site and the 500 bp on either side of the middle of the peak (Figure 4A). However, conservation decreased to background levels at a distance of 1 kb or more from the center of the ER binding sites. This supports the hypothesis that the discrete ER binding sites we see in MCF-7 cells are conserved between species and likely play a more general role in ER action in other cellular systems.

A Screen for Common Sequences Enriched in Genuine ER Binding Regions Suggests the Importance of Forkhead Factors in Estrogen Action

An unbiased search for common sequence motifs (Liu et al., 2002) within the 57 individual ER binding sites on chromosomes 21 and 22 revealed the significant recurrence of two motifs. A consensus 15 base sequence identical to the canonical ERE was present in 49% of all the ER binding sites on chromosomes 21 and 22 (Figure 4B; Klinge, 2001). The likelihood of an ERE occurring in one of the ER binding sites was significantly increased when compared to all of chromosomes 21 and 22 ($p = 1.33E-15$). In the ER binding sites lacking a canonical ERE, a majority were found to contain one or more ERE half-sites, and the occurrence of ERE half-sites was also nonrandom ($p = 2.16E-14$). To confirm that our failure to find ER binding at other EREs (5500 predicted EREs on chromosomes 21 and 22, as listed in Figures 1A and 1B) was not due to the insensitivity of our ChIP-microarray technique, we performed ChIP for ER followed by PCR for several randomly selected, predicted but nonfunctional perfect EREs on chromosomes 21 and 22. No ER association was found at any of these sites (data not shown).

We next determined whether DNA sequences other than the classical ERE were found at the ER binding sites by analyzing the bound sequences for conserved motifs after removing the EREs. This analysis revealed the presence of a Forkhead factor binding site in 54% of the 57 ER binding regions (Figure 4B), a finding that would only occur by chance with a probability of $p = 1.23E-8$. Forkhead binding motifs were found in 56% of the ER binding regions that contain a canonical ERE. Using the consensus Forkhead motif recurring within these regions (Figure 4B), we determined the probability of this motif residing within predicted ERE regions that are not bound by ER in vivo (18.45%). This significant enrichment of a Forkhead motif within ER binding regions ($p = 3.78E-7$) suggested the presence of adjacent Forkhead motifs may play a role in determining ER binding. The finding that the largest category of sites contains both an ERE and a Forkhead motif (47.4%) strongly suggests a functional interaction (Figure 4C).

Forkhead Proteins Play a Combinatorial and Essential Role in ER Binding and ER-Mediated Gene Transcription

A combinatorial interaction between Forkhead and ER pathways has been previously suggested for a small number of specific genes. HNF-3α (FoxA1) Forkhead binding domains within the promoter of the estrogen-regulated genes *TFF-1* (Beck et al., 1999) and Vitellogenin B1 (Robyr et al., 2000) have been shown to be important for gene transcription, and they have been shown to interact directly with ER in yeast two-hybrid experiments (Schuur et al., 2001). The function of Forkhead proteins can be regulated by their nuclear-cytoplasmic distribution depending on their phosphorylation (Brunet et al., 1999; Kops et al., 1999). We therefore determined that FoxA1 localized to the nucleus before and after estrogen stimulation of MCF-7 cells (data not shown).

We next determined whether FoxA1 was recruited along with ER to the ER binding domains. Directed ChIP of FoxA1 followed by real-time PCR of all 57 ER binding regions on chromosomes 21 and 22 revealed a high degree of concordance between regions that recruit ER and FoxA1. Approximately 48% of all of the ER

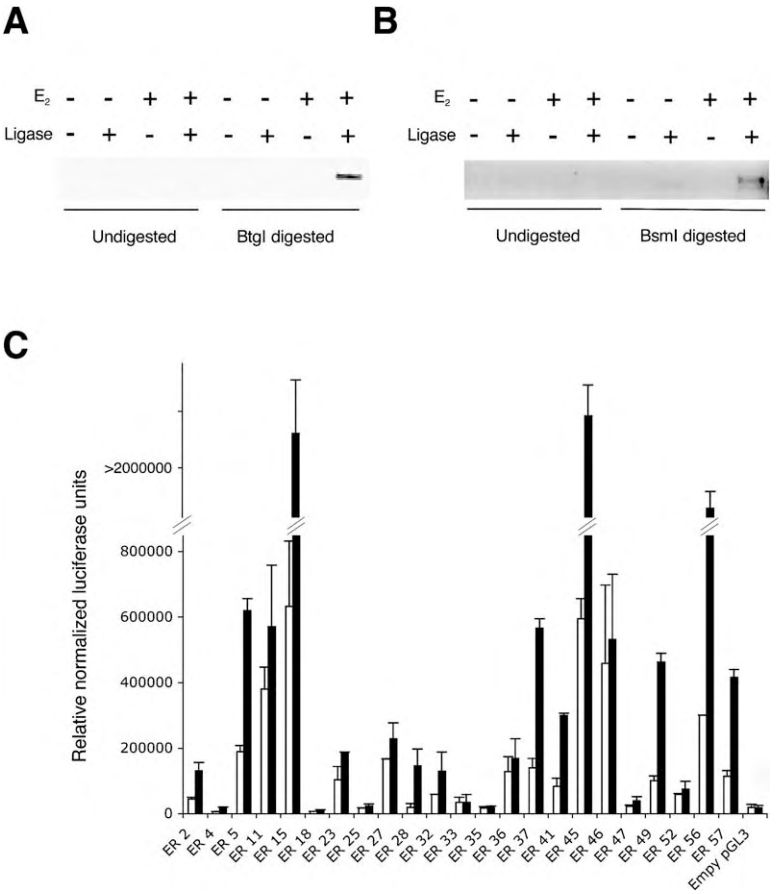


Figure 3. Interaction of Promoter-Enhancer Domains and Transcriptional Activity of Enhancer Regions

(A) Chromosome capture assay was performed after digesting fixed chromatin from vehicle- or estrogen-treated cells with the BtgI restriction enzyme. Primers flanking the *TFF-1* promoter and enhancer were used to amplify DNA after ligation. Undigested controls and no ligase controls were included.

(B) Chromatin was digested with BsmI, and one primer flanking the *NRIP-1* promoter and one in enhancer 3 region were used to amplify a specific product after ligation.

(C) ER binding sites were cloned into the pGL-3 promoter vector and transfected into hormone-depleted MCF-7 cells, after which vehicle (open bars) or estrogen (solid bars) was added. Empty pGL3-promoter vector was used as a negative control. Cotransfection of pRL null Renilla vector was included as a normalizing control. The data are the average of three replicates ± SD.

binding domains showed FoxA1 interaction, although the pattern of recruitment differed from site to site (Figure S3). A majority of the regions containing FoxA1 did so in the absence of estrogen, but FoxA1 binding was decreased following estrogen stimulation. This was the case for *NRIP-1* enhancer 1, *DSCAM-1* enhancer 1, and *TFF-1* promoter (Figure 5A). FoxA1 association with *XBP-1* enhancer 2 was clearly observed but was not diminished after estrogen addition (Figure 5A). All of these ER binding sites contained a Forkhead motif and an ERE or ERE half-site (Figure 5B). FoxA1 was not seen to bind to *XBP-1* enhancer 3, which lacks a Forkhead motif (Figure 5). However, several regions containing Forkhead motifs did not recruit FoxA1, and several ER binding domains that lacked Forkhead motifs did bind FoxA1. This complex interplay between FoxA1, ER, and binding sites within chromatin likely involves adjacent regions to the ER binding sites and may involve other proteins. Despite this, it is clear that a significant proportion of ER binding sites, especially those adjacent to actively transcribed genes, contain FoxA1 prior to estrogen stimulation and ER recruitment to the same regions.

To determine the importance of FoxA1 in mediating ER association with chromatin, we developed siRNA to the 3′UTR of *FoxA1* mRNA. Specific targeted knockdown of FoxA1 protein was achieved (Figure 6A), without changes in control protein or ER protein levels (data not shown). A luciferase siRNA (siLuc) was used as a negative control. MCF-7 cells were deprived of hormones for 24 hr and siLuc, or siRNA to *FoxA1*, was transfected for 6 hr, after which hormone-depleted media was added for a further 48 hr and cells were stimulated with estrogen or vehicle. ER ChIP and real-time PCR of a number of previously validated binding sites was performed. The decrease in *FoxA1* completely impeded the ability of ER to bind to *TFF-1* promoter, *XBP-1* enhancer 1, and *NRIP-1* enhancer 2 (Figure 6B), as well as *DSCAM-1* enhancer 1 (data not shown). No changes were observed on the *XBP-1* promoter, which functioned as a negative control (Figure 6B).

Since the targeted knockdown of FoxA1 inhibited the ability of ER to associate with in vivo ER binding sites, we assessed the effect of Forkhead downregulation on estrogen-mediated transcription. After siLuc or siFoxA1 transfection, cells were stimulated with estrogen or vehicle for 6 hr and mRNA changes in all 12 estrogen target genes on chromosomes 21 and 22 were assessed. The estrogen-induced increases in all 12 estrogen targets were abolished when FoxA1 was downregulated (Figure 6C), but no changes were observed in *GAPDH* control mRNA levels. The essential role for the FoxA1 Forkhead protein during transcription of all estrogen target genes on chromosomes 21 and 22 confirms a general requirement of FoxA1 for ER transcription.

Discussion

A complete picture of ER-mediated gene activation has begun to emerge in recent years, with a coordinated

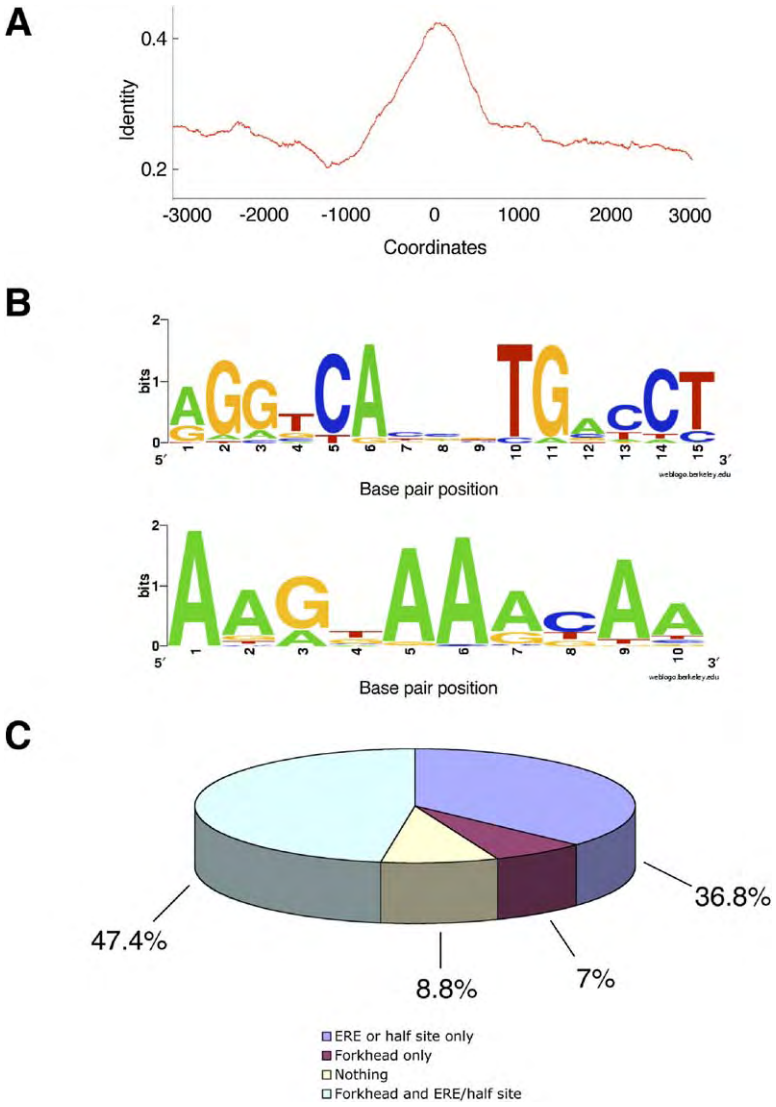


Figure 4. Conservation of ER Binding Sites and Presence of Enriched Motifs

(A) Sequence homology of ER binding sites and surrounding sequence between human and mouse genomes. The center of ER peaks is designated coordinate 0.

(B) An unbiased motif screen of all the ER binding sites on chromosomes 21 and 22 revealed the presence of two enriched motifs, an ERE and a Forkhead binding motif, both of which are visually represented in WebLogo (http://weblogo.berkeley.edu).

(C) The occurrence of ERE or ERE half-sites and Forkhead sites within the 57 ER binding sites on chromosomes 21 and 22.

and timely cycling of receptor, nuclear coactivators, chromatin remodelling proteins, and the transcription machinery on and off target promoters (Metivier et al., 2003; Shang et al., 2000). However, these studies oversimplify the problem by focusing on the promoter proximal region of one or two target genes and largely ignore the remaining chromosomal sequence. Here, we have interrogated the association of ER across entire chromosomes, including intergenic regions that contain potential *cis*-regulatory domains. These ChIP-microarray experiments demonstrate the ability to identify genuine in vivo ER protein binding sites in previously unexplored regions of the genome. Interestingly, while a few of the ER binding sites were found directly adjacent to ER target genes, most were found at significant distances including several >100 kb removed from transcription start sites. Of the 57 ER binding sites (within 32 potential transcriptional regulatory clusters), only a very small number of proximal promoters recruited ER, despite the fact that the other genes were estrogen induced. The presence of multiple components of the transcriptional machinery at distal sites combined with the ability of chromosome conformation capture assays to demonstrate that these distant sites are physically associated with promoter-proximal regions suggests that they play an important role in estrogen-mediated regulation.

A significant volume of work has focused on identifying essential domains within the proximal promoters of known estrogen regulated genes (Dubik and Shiu, 1992; Petz et al., 2002; Porter et al., 1996; Teng et al., 1992; Umayahara et al., 1994; Vyhlidal et al., 2000; Weisz and Rosales, 1990). The conclusions drawn from this large volume of data implicate a number of motifs, including Sp1, AP-1, and GC-rich regions as important *cis*-regulatory domains in ER-mediated transcription. However, our data demonstrate ER regulatory sites at distances several orders of magnitude greater than was focused on in the past, suggesting that they may function in ways analogous to the β-*globin* LCR (Sawado et al., 2003).

Nonbiased motif scanning of the genuine in vivo ER binding sites identified a canonical ERE in the majority of ER binding sites that represented only 1.5% of EREs

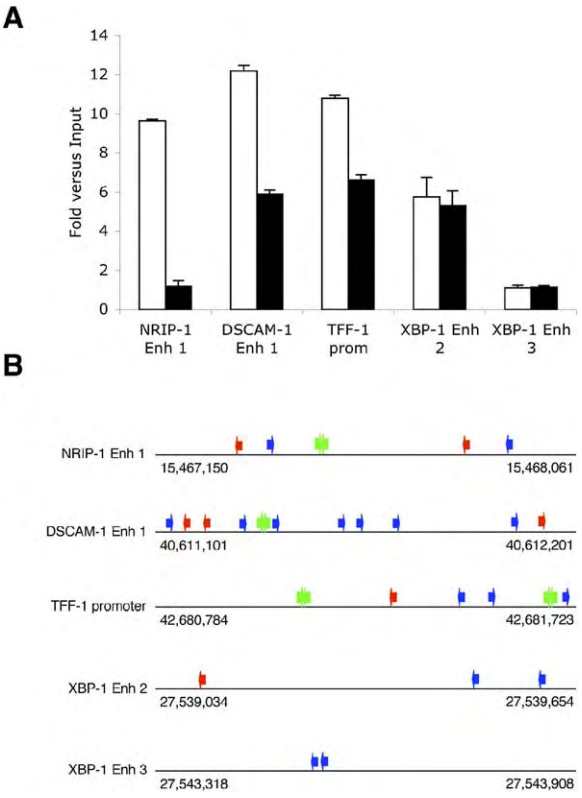


Figure 5. Recruitment of Forkhead Protein to ER Binding Domains

(A) ChIP of FoxA1 followed by real-time PCR of *NRIP-1* enhancer 1, *DSCAM-1* enhancer 1, *TFF-1* promoter, and *XBP-1* enhancer 2. *XBP-1* enhancer 3 is included as a control which does not recruit FoxA1. Data are shown as fold change versus input and are the average of three replicates ± SD. Open bars are vehicle treated and solid bars are estrogen treated.

(B) Schematic diagram showing the relative location of ERE motifs (inverted green arrows), ERE half-sites (blue arrows), and Forkhead motifs (red arrows). Chromosome nucleotide locations are given.

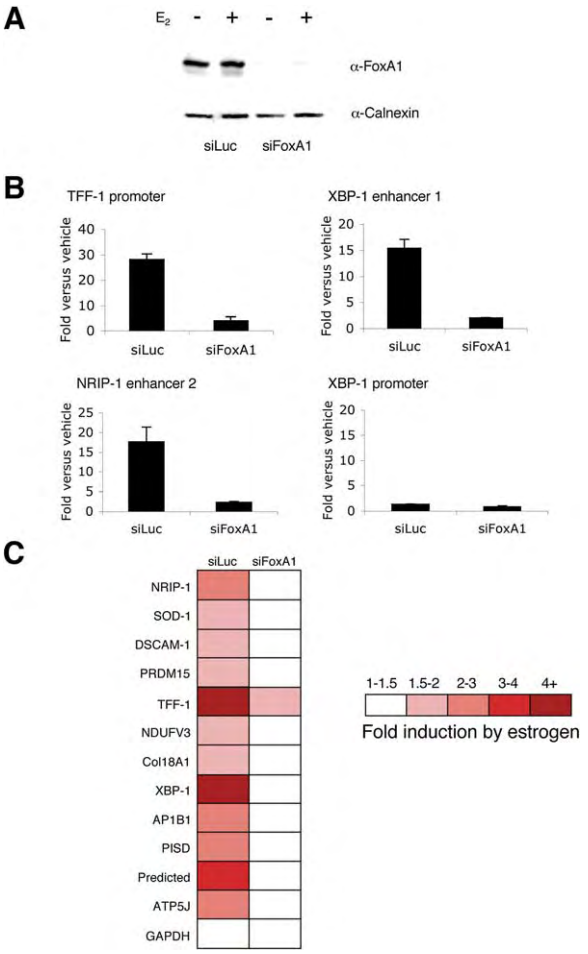


Figure 6. Specific Targeted Knockdown of FoxA1 and the Effects on Estrogen-Mediated Transcription

(A) siRNA to FoxA1 was transfected into hormone-depleted MCF-7 cells, and changes in FoxA1 protein levels were determined after vehicle or estrogen treatment. SiLuc was used as a transfection control and Calnexin was used as a loading control.

(B) ER ChIP was performed after vehicle or estrogen treatment of siLuc or siFoxA1 transfected cells and real-time PCR was conducted on *TFF-1* promoter, *XBP-1* enhancer 1, *NRIP-1* enhancer 2, as well as *XBP-1* promoter as a negative control. The data are fold enrichment over vehicle-treated.

(C) Changes in mRNA levels of all estrogen-regulated genes on chromosomes 21 and 22 after siLuc or siFoxA1. The data are estrogen-mediated fold enrichment compared to vehicle (ethanol) control and are the average of three separate replicates ± SD. The color intensity reflects the fold change as described in the legend.

predicted by bioinformatics alone. Previous approaches for motif identification involved computational-based methods for identifying response elements, after which gene proximal sites are included as potential binding domains (Bajic and Seah, 2003; Bourdeau et al., 2004). The current data suggest that while ER binding involves interaction with consensus ERE motifs, the presence of such motifs is insufficient to dictate receptor-chromatin association. Furthermore, the exclusion of response elements further than several kilobases from transcription start sites eliminates distal regulatory regions that may be the primary receptor-chromatin interaction sites.

Since the presence of an ERE alone is insufficient to define an authentic ER regulatory site, we searched for other conserved sequences and found that Forkhead factor binding sites are present near authentic EREs significantly more frequently than those that do not bind ER. We showed that a Forkhead factor (FoxA1) binding was essential for ER-chromatin interactions and subsequent expression of estrogen gene targets. A link between *ER* and *FoxA1* has previously been shown, with their expression correlated in breast cancer cell lines (Lacroix and Leclercq, 2004). FoxA1 protein can bind condensed chromatin via its winged-helix DNA binding domains that mimic histone linker proteins (Cirillo et al., 2002; Cirillo et al., 1998). Unlike histone proteins however, FoxA1 does not contain the amino acid composition to condense chromatin and it therefore is thought to promote euchromatic conditions. As such, it is possible that the presence of FoxA1 identifies specific regions within chromatin to facilitate the association of the ER transcription complex. Our data suggest that FoxA1 is present on the chromatin at a number of regions, after which ER can associate with these spe-

cific sites. Downregulation of FoxA1 inhibits the ability of ER to associate with its binding sites, confirming the requirement for Forkhead-directed association of ER with chromatin, despite the fact that these sites contain sufficient information, in the form of an ERE, for ER docking. This, combined with a recent investigation showing that FoxA1 can directly modulate chromatin in the MMTV promoter and can positively enhance transcription by the glucocorticoid receptor (Holmqvist et al., 2005), supports a general model for FoxA1 involvement in nuclear receptor transcription.

We have taken an unbiased approach to identify regions of chromatin, both promoter proximal and intergenic sequences, which are involved in ER-mediated transcriptional activity. We find a limited number of bona fide ER binding sites on chromosomes 21 and 22, with a significant enrichment of canonical ERE palindromes and half-sites within the binding sites. Moreover, the presence of Forkhead binding motifs and the subsequent identification of a functional role for the Forkhead protein FoxA1 in estrogen signaling exemplifies the power of this approach to identify important regulatory domains within the vast regions of unexplored sequence of the human genome.

Experimental Procedures

Chromatin Immunoprecipitation (ChIP)-Microarray Preparation

ChIP was performed as previously described (Shang et al., 2000), with the following modifications. Two micrograms of antibody was prebound for a minimum of 4 hr to protein A and protein G Dynal magnetic beads (Dynal Biotech, Norway) and washed three times with ice-cold PBS plus 5% BSA and then added to the diluted chromatin and immunoprecipitated overnight. The magnetic bead-chromatin complexes were collected and washed six times in RIPA buffer (50 mM HEPES [pH 7.6], 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5 M LiCl). Elution of the DNA from the beads was as previously described (Shang et al., 2000). Antibodies used were as follows: ERα (Ab-10) from Neomarkers (Lab Vision, United Kingdom), ERα (HC-20), RNA PolII (H-224), AIB-1/RAC3 (C-20), HNF-3α/FoxA1 (H-120), mouse IgG (sc-2025), and rabbit IgG (sc-2027) from Santa Cruz (Santa Cruz Biotechnologies, California). Ligation-Mediated PCR was performed as previously described (Ren et al., 2002). Labeling was performed as previously described (Kapranov et al., 2002). Microarrays used were Affymetrix Genechip chromosome 21/22 tiling set P/N 900545.

Data Analysis

1,054,325 probe pairs were mapped to chromosomes 21 and 22 according to the NCBIv33 GTRANS Libraries provided by Affymetrix. (PM-MM) value was recorded for each probe pair, and a probe pair was removed if either PM or MM was flagged as outlier by the Affymetrix GCOS software. The samples (three ER+ ChIP and three genomic inputs) were normalized by quantile normalization (Bolstad et al., 2003) based on a combined 76 ChIP experiments obtained from public domain and Dana-Farber Cancer Institute. The behavior of every probe pair i, assumed to be $N(\mu_i, \sigma_i^2)$, was estimated from the 76 normalized experiments. A two-state (ChIP-enriched state and nonenriched state) Hidden Markov Model with the following parameters was applied to each sample to estimate the probability of ChIP enrichment at each probe pair location:

Transition probabilities: 300/1,054,325 for transition to a different state,
1 – (300/1,054,325) for staying in the same state.
Emission probabilities: $N(\mu_i, \sigma_i^2)$ for nonenriched hidden state,
$N(\mu_i + 2\sigma_i, (1.5\sigma_i)^2)$ for enriched hidden state.

To combine the results from the six samples, an enrichment score was calculated as the average enrichment probability in the three ER+ ChIP samples subtracted by the average enrichment probability in the three genomic input samples. Since the tiling array has one 25-mer probe in every 35 bp of nonrepeat regions, the coverage of every probe was extended by 10 bp on both ends. An enriched regions is defined as run of probes with enrichment score >50% and covering at least 125 bp. Each enriched region can tolerate up to two neighboring probes with enrichment score between [10%, 50%]. If two neighboring probes are more than 210 bp apart, the enriched region is broken into two separate blocks. A summary enrichment score was obtain for each enriched region, which is the enrichment score summation for all the probes in the region divided by the square root of the number of probes in the region. This summary enrichment score represents the relative confidence of a predicted enriched region.

Sequence Analysis

The genomic DNA of every ChIP-enriched region was retrieved from UCSC genome browser and ranked by the summary enrichment score. MDscan algorithm (Liu et al., 2002) was applied to the sequences to find enriched sequence pattern that is the putative estrogen receptor binding motif. To find a motif of width w, MDscan first enumerates each w-mer in the highest ranking sequences and collects other w-mers similar to it in these sequences to construct a candidate motif as a probability matrix. A semi-Bayes scoring function was used to remove low-scoring candidate motifs and refine the rest by checking all w-mers in all the ChIP-enriched sequences. A high-scoring motif (with similar consensus) consistently reported multiple times at different motif widths indicates a strong prediction.

We expanded all 57 of the ER binding sites equally in each direction to have a length of 6 kb. The human-mouse conservation score of each nucleotide in the expanded binding region is defined as the average sequence identity (# matched nucleotides – # indels)/500 of a 500-mer window centered at the nucleotide. The human (hg15) /mouse (mm3) BLASTZ (Schwartz et al., 2003) genome alignments were downloaded from http://genome.ucsc.edu.

Real-Time PCR

Primers were selected using Primer Express (Applied Biosystems). Five microliters of precipitated and purified DNA was subjected to PCR using the Applied Biosystems SYBR Green Mastermix. Relative DNA quantities were measured using the PicoGreen system (Molecular Probes, Oregon). All primer sequences and locations are listed in Table S2.

Double-Stranded cDNA Synthesis

Total RNA was converted to double stranded cDNA according to the Invitrogen Superscript double-stranded cDNA synthesis manufacturer's instructions. The RNA was primed with 250 ng oligo(dT) (Invitrogen) and 25 ng random hexamers (Gibco). cDNA was fragmented and labeled as described above.

5′RACE

5′ RACE was performed according to the manufacturer's instructions (Invitrogen). The primers sequences used were as follows: *NRIP-1* RT primer (5′-TGCCTGATGCATTAGTAATCC-3′), *NRIP-1* nested primer 1 (5′-GAGCCAAGCTCTTCTCCATGAGTCATGTTC-3′), and *NRIP-1* nested primer 2 (5′-ACCTTCCATCGCAATCAGAGA GAGACGTACTG-3′). The PCR product was cloned and sequenced by standard methods.

Chromosome Capture Assay

Fixed chromatin was digested overnight with specific restriction enzymes after which ER ChIP was set up as described above. After overnight ChIP, the beads were precipitated and resuspended in ligation buffer (NEB, Massachusetts) and overnight ligation was performed. The beads were collected, washed, and the formaldehyde crosslinking was reversed as described above. Primers used to amplify annealed fragments were as described in Table S2.

Luciferase Enhancer Activity

ER binding sites were amplified by PCR and cloned into the pGL-3-promoter vector (Promega). Hormone-depleted MCF-7 cells were transfected with each of the ER binding domain vectors with Lipofectamine 2000 (Invitrogen), and total protein lysate was harvested after estrogen or ethanol addition for 24 hr. Transfections were normalized by the cotransfection of the pRL null renilla luciferase vector and renilla and firefly luciferase activity was assessed using the dual luciferase kit (Promega).

Western Blotting

SDS-PAGE was performed as previously described (Carroll et al., 2000). Antibodies used were FoxA1/HNF-3α (ab5089), from AbCam (Cambridge, United Kingdom) and Calnexin (H-70) from Santa Cruz (California).

Short Interfering (si) RNA

A 21 bp siRNA was designed against the FoxA1 transcript and synthesized by Dharmacon (Lafayette, Colorado). siRNA was transfected using Lipofectamine 2000 (Invitrogen). The siRNA sequences used were as follows: *siFoxA1* sense 5′-GAGAGAAAAAA UCAACAGC-3′ and antisense 5′-GCUGUUGAUUUUUUCUCUC-3′; siLuc sense 5′-CACUUACGCUGAGUACUUCGA-3′ and antisense 5′-UCGAAGUACUCAGCGUAAGUG-3′.

Supplemental Data

Supplemental Data include four figures, two tables, and raw data files and can be found with this article online at http://www.cell.com/cgi/content/full/122/1/33/DC1/.

Acknowledgments

The authors would like to thank Danielle Lynch and Mitch Lazar for helpful discussions and review of the manuscript. We thank Molin Wang, David Harrington, Cheng Li, and Jun S. Liu for their insights and advice on data analysis. This work was supported by grants from the Claudia Adams Barr Program in Innovative Basic Cancer Research (M.B., P.A.S., X.S.L.), the DF/HCC Breast Cancer SPORE Grant, by DOD Breast Cancer Research Program Awards DAMD17-01-1-0222 (M.B.), W81XWH-04-1-0512 (J.S.C.), DAMD17-02-0364 (P.A.S.), by N.I.H. grants K22 HG02488-01A1 (A.S.B.), F32 CA108380-02 (T.R.G.), T90 DK070078-01 (C.A.M.), and by fellowships from the Fondation Recherche Medicale (J.E.), and the HHMI (A.J.S.).

Received: January 27, 2005
Revised: April 6, 2005
Accepted: May 6, 2005
Published: July 14, 2005

References

Augereau, P., Miralles, F., Cavailles, V., Gaudelet, C., Parker, M., and Rochefort, H. (1994). Characterization of the proximal estrogen-responsive element of human cathepsin D gene. Mol. Endocrinol. *8*, 693–703.

Bajic, V.B., and Seah, S.H. (2003). Dragon gene start finder identifies approximate locations of the 5′ ends of genes. Nucleic Acids Res. *31*, 3560–3563.

Beck, S., Sommer, P., dos Santos Silva, E., Blin, N., and Gott, P. (1999). Hepatocyte nuclear factor 3 (winged helix domain) activates trefoil factor gene TFF1 through a binding motif adjacent to the TATAA box. DNA Cell Biol. *18*, 157–164.

Beckman, J.S., Carson, M., Smith, C.D., and Koppenol, W.H. (1993). ALS, SOD and peroxynitrite. Nature *364*, 584.

Berry, M., Nunez, A.M., and Chambon, P. (1989). Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc. Natl. Acad. Sci. USA *86*, 1218–1222.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Bourdeau, V., Deschenes, J., Metivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J.H., and Mader, S. (2004). Genome-wide identification of high-affinity estrogen response elements in human and mouse. Mol. Endocrinol. *18*, 1411–1427.

Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J., and Greenberg, M.E. (1999). Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. Cell *96*, 857–868.

Bulger, M., Sawado, T., Schubeler, D., and Groudine, M. (2002). ChIPs of the beta-globin locus: unraveling gene regulation within an active domain. Curr. Opin. Genet. Dev. *12*, 170–177.

Carroll, J.S., Prall, O.W., Musgrove, E.A., and Sutherland, R.L. (2000). A pure estrogen antagonist inhibits cyclin E-Cdk2 activity in MCF-7 breast cancer cells and induces accumulation of p130–E2F4 complexes characteristic of quiescence. J. Biol. Chem. *275*, 38221–38229.

Cavailles, V., Dauvois, S., L'Horset, F., Lopez, G., Hoare, S., Kushner, P.J., and Parker, M.G. (1995). Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. EMBO J. *14*, 3741–3751.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499–509.

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., and Zaret, K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244–254.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Dubik, D., and Shiu, R.P. (1992). Mechanism of estrogen activation of c-myc oncogene expression. Oncogene *7*, 1587–1594.

Giamarchi, C., Solanas, M., Chailleux, C., Augereau, P., Vignon, F., Rochefort, H., and Richard-Foy, H. (1999). Chromatin structure of the regulatory regions of pS2 and cathepsin D genes in hormone-dependent and -independent breast cancer cell lines. Oncogene *18*, 533–541.

Halachmi, S., Marden, E., Martin, G., MacKay, H., Abbondanza, C., and Brown, M. (1994). Estrogen receptor-associated proteins: possible mediators of hormone-induced transcription. Science *264*, 1455–1458.

Holmqvist, P.-H., Belikov, S., Zaret, K.S., and Wrange, Ö. (2005). FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. Exp. Cell Res. *304*, 593–603.

Horike, S., Cai, S., Miyano, M., Cheng, J.F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat. Genet. *37*, 31–40.

Ikeda, K., Sato, M., Tsutsumi, O., Tsuchiya, F., Tsuneizumi, M., Emi, M., Imoto, I., Inazawa, J., Muramatsu, M., and Inoue, S. (2000). Promoter analysis and chromosomal mapping of human EBAG9 gene. Biochem. Biophys. Res. Commun. *273*, 654–660.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science *296*, 916–919.

Katsanis, N., Ives, J.H., Groet, J., Nizetic, D., and Fisher, E.M. (1998). Localisation of receptor interacting protein 140 (RIP140) within 100 kb of D21S13 on 21q11, a gene-poor region of the human genome. Hum. Genet. *102*, 221–223.

Klinge, C.M. (2001). Estrogen receptor interaction with estrogen response elements. Nucleic Acids Res. *29*, 2905–2919.

Kops, G.J., de Ruiter, N.D., De Vries-Smits, A.M., Powell, D.R., Bos,

J.L., and Burgering, B.M. (1999). Direct control of the Forkhead transcription factor AFX by protein kinase B. Nature *398*, 630–634.

Kuang, S.Q., Liao, L., Zhang, H., Lee, A.V., O'Malley, B.W., and Xu, J. (2004). AIB1/SRC-3 deficiency affects insulin-like growth factor I signaling pathway and suppresses v-Ha-ras-induced breast cancer initiation and progression in mice. Cancer Res. *64*, 1875–1885.

Lacroix, M., and Leclercq, G. (2004). About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. Mol. Cell. Endocrinol. *219*, 1–7.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat. Biotechnol. *20*, 835–839.

Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751–763.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378–1381.

Pedram, A., Razandi, M., Aitkenhead, M., Hughes, C.C., and Levin, E.R. (2002). Integration of the non-genomic and genomic actions of estrogen. Membrane-initiated signaling by steroid to transcription and cell biology. J. Biol. Chem. *277*, 50768–50775.

Petz, L.N., Ziegler, Y.S., Loven, M.A., and Nardulli, A.M. (2002). Estrogen receptor alpha and activating protein-1 mediate estrogen responsiveness of the progesterone receptor gene in MCF-7 breast cancer cells. Endocrinology *143*, 4583–4591.

Porter, W., Wang, F., Wang, W., Duan, R., and Safe, S. (1996). Role of estrogen receptor/Sp1 complexes in estrogen-induced heat shock protein 27 gene expression. Mol. Endocrinol. *10*, 1371–1378.

Rabiner, L.R. (1989). A tutorial on Hidden Markov models and selected applications in speech recognition. Proc. IEEE *77*, 257–286.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev. *16*, 245–256.

Robyr, D., Gegonne, A., Wolffe, A.P., and Wahli, W. (2000). Determinants of vitellogenin B1 promoter architecture. HNF3 and estrogen responsive transcription within chromatin. J. Biol. Chem. *275*, 28291–28300.

Sawado, T., Halow, J., Bender, M.A., and Groudine, M. (2003). The beta-globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. Genes Dev. *17*, 1009–1018.

Schiff, R., Reddy, P., Ahotupa, M., Coronado-Heinsohn, E., Grim, M., Hilsenbeck, S.G., Lawrence, R., Deneke, S., Herrera, R., Chamness, G.C., et al. (2000). Oxidative stress and AP-1 activity in tamoxifen-resistant breast tumors in vivo. J. Natl. Cancer Inst. *92*, 1926–1934.

Schuur, E.R., Loktev, A.V., Sharma, M., Sun, Z., Roth, R.A., and Weigel, R.J. (2001). Ligand-dependent interaction of estrogen receptor-alpha with members of the forkhead transcription factor family. J. Biol. Chem. *276*, 33554–33560.

Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res. *13*, 103–107.

Shang, Y., and Brown, M. (2002). Molecular determinants for the tissue specificity of SERMs. Science *295*, 2465–2468.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A., and Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell *103*, 843–852.

Singh, R.J., Karoui, H., Gunther, M.R., Beckman, J.S., Mason, R.P., and Kalyanaraman, B. (1998). Reexamination of the mechanism of hydroxyl radical adducts formed from the reaction between familial amyotrophic lateral sclerosis-associated Cu,Zn superoxide dismutase mutants and H2O2. Proc. Natl. Acad. Sci. USA *95*, 6675–6680.

Teng, C.T., Liu, Y., Yang, N., Walmer, D., and Panella, T. (1992). Differential molecular mechanism of the estrogen action that regulates lactoferrin gene in human and mouse. Mol. Endocrinol. *6*, 1969–1981.

Torres-Arzayus, M.I., De Mora, J.F., Yuan, J., Vazquez, F., Bronson, R., Rue, M., Sellers, W.R., and Brown, M. (2004). High tumor incidence and activation of the PI3K/AKT pathway in transgenic mice define AIB1 as an oncogene. Cancer Cell *6*, 263–274.

Umayahara, Y., Kawamori, R., Watada, H., Imano, E., Iwama, N., Morishima, T., Yamasaki, Y., Kajimoto, Y., and Kamada, T. (1994). Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. J. Biol. Chem. *269*, 16433–16442.

Vyhlidal, C., Samudio, I., Kladde, M.P., and Safe, S. (2000). Transcriptional activation of transforming growth factor alpha by estradiol: requirement for both a GC-rich site and an estrogen response element half-site. J. Mol. Endocrinol. *24*, 329–338.

Wang, D.Y., Fulthorpe, R., Liss, S.N., and Edwards, E.A. (2004). Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. Mol. Endocrinol. *18*, 402–411.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev. *16*, 235–244.

Weisz, A., and Rosales, R. (1990). Identification of an estrogen response element upstream of the human c-fos gene that binds the estrogen receptor and the AP-1 transcription factor. Nucleic Acids Res. *18*, 5097–5106.

nature genetics

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Genome-wide analysis of estrogen receptor binding sites

Jason S Carroll[1], Clifford A Meyer[2,3], Jun Song[2,3], Wei Li[2,3], Timothy R Geistlinger[1], Jérôme Eeckhoute[1], Alexander S Brodsky[4], Erika Krasnickas Keeton[1], Kirsten C Fertuck[1], Giles F Hall[5], Qianben Wang[1], Stefan Bekiranov[6,8], Victor Sementchenko[6], Edward A Fox[5], Pamela A Silver[5,7], Thomas R Gingeras[6], X Shirley Liu[2,3] & Myles Brown[1]

The estrogen receptor is the master transcriptional regulator of breast cancer phenotype and the archetype of a molecular therapeutic target. We mapped all estrogen receptor and RNA polymerase II binding sites on a genome-wide scale, identifying the authentic *cis* binding sites and target genes, in breast cancer cells. Combining this unique resource with gene expression data demonstrates distinct temporal mechanisms of estrogen-mediated gene regulation, particularly in the case of estrogen-suppressed genes. Furthermore, this resource has allowed the identification of *cis*-regulatory sites in previously unexplored regions of the genome and the cooperating transcription factors underlying estrogen signaling in breast cancer.

Recent work has focused on identifying gene expression signatures in breast cancer subtypes that predict response to specific treatment regimes and improved disease outcome[1–4]. Tumors with gene expression profiles that correlate with estrogen receptor α (hereafter referred to simply as 'estrogen receptor') expression have been termed luminal type[1,5] and are associated with response to endocrine therapy and improved survival, although the mechanisms by which estrogen receptor dictates tumor status are poorly understood.

Estrogen receptor–mediated transcription has been intensively studied on a small number of endogenous target promoters[6–8], and recent location analysis of estrogen receptor binding by chromatin immunoprecipitation combined with microarrays (ChIP-on-chip) also focused primarily on promoter regions[9,10]. We recently expanded on these analyses to map estrogen receptor binding sites in a less biased way that did not depend on preexisting concepts of classic promoter domains[11] and subsequently identified several new features of estrogen receptor transcription, including an involvement of distal *cis*-regulatory enhancer regions and a requirement for the Forkhead protein, FoxA1, in facilitating estrogen receptor binding to chromatin and subsequent gene transcription[11]. This analysis highlighted the importance of regions of chromatin distinct from the promoter-proximal regions and suggested an *in vivo* requirement for cooperating transcription factors. However, owing to technological limitations, this investigation was restricted to chromosomes 21 and 22, comprising <3% of the genome and containing few estrogen receptor–regulated genes[11]. Recent chromosome-wide transcript analyses have demonstrated the existence of multiple layers of transcription that are independent of known coding gene regions[12], implying that transcription factor activity cannot be described by a limited set of paradigms that are restricted to well-studied regions of the genome. To overcome these limitations, we conducted a genome-wide analysis of estrogen receptor and RNA polymerase II (PolII) binding by mapping estrogen-induced estrogen receptor and RNA PolII binding sites on all 1,500 Mb of nonrepetitive sequence in a breast cancer cell line at 35-bp resolution. The combination of this unique resource with gene expression data serves to elucidate the mechanisms underlying estrogen-regulated gene expression in breast cancer.

RESULTS

The MCF-7 breast cancer cell line has been extensively used as a model of hormone-dependent breast cancer. We deprived MCF-7 cells of hormones for 3 d and then synchronously induced transcription by the addition of estrogen for a brief period of time (45 min) known to result in maximal estrogen receptor–chromatin binding[6,11]. We used estrogen receptor–specific and RNA PolII–specific antibodies for ChIP and prepared precipitated chromatin as previously described[11]. We hybridized ChIP chromatin and input DNA to the Affymetrix Human tiling 1.0 microarrays representing the entire nonrepetitive human genome sequence (NCBI build 35) tiled at 35-bp resolution. We performed three biological replicates and identified enriched binding sites (**Supplementary Note** online) by the intersection of two independent methods: namely, a nonparametric generalized

[1]Department of Medical Oncology and [2]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney St., Boston, Massachusetts 02115, USA. [3]Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA. [4]Brown University, Laboratories for Molecular Medicine, Center for Genomics and Proteomics, Providence, Rhode Island 02903, USA. [5]Department of Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney St., Boston, Massachusetts 02115, USA. [6]Affymetrix, 3380 Central Expressway, Santa Clara, California 95051, USA. [7]Department of Systems Biology, Harvard Medical School, 200 Longwood Ave., Boston, Massachusetts 02115, USA. [8]Current address: Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA. Correspondence should be addressed to M.B. (myles_brown@dfci.harvard.edu) or X.S.L. (xsliu@jimmy.harvard.edu).

Received 19 July; accepted 8 September; published online 1 October 2006; corrected online 20 October 2006 (details online); doi:10.1038/ng1901

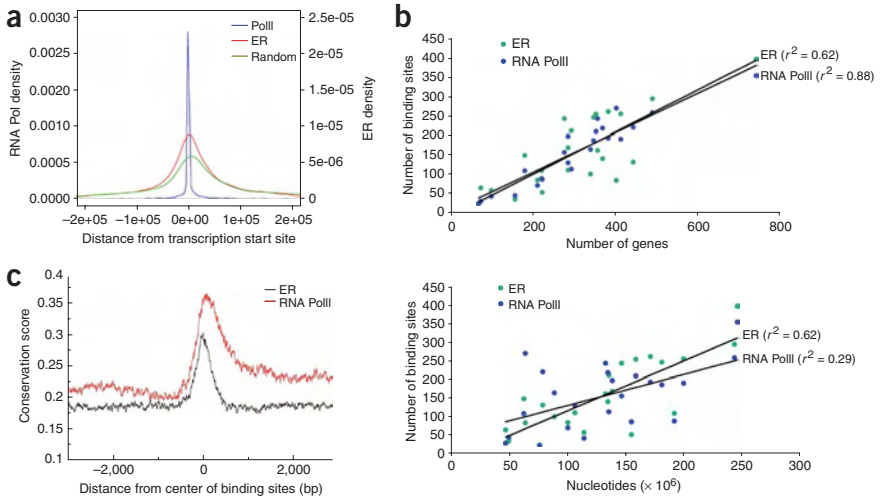


Figure 1 Summary of estrogen receptor and RNA PolII binding sites and correlation with nucleotide and gene number. (**a**) Location of estrogen receptor (ER) and RNA PolII sites relative to transcription start sites (TSS) of RefSeq genes. The scale on the left represents RNA PolII distribution, and the scale on the right represents estrogen receptor and random distribution. (**b**) Correlation of estrogen receptor and RNA PolII binding sites with each chromosome, ranked according to total gene number and total nucleotide number. (**c**) Conservation of all estrogen receptor binding sites (black line) and RNA PolII binding sites (red line) between human, mouse, rat, chicken and *Fugu rubripes* sequence. RNA PolII binding sites are shown in a 5′-to-3′ manner.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Mann-Whitney U-test[13] using a threshold of $P < 10^{-5}$ and a new model-based analysis tiling array algorithm, MAT[14]. This stringent approach ensures high confidence predictions to facilitate subsequent motif analysis, though it may introduce some false negatives with lower confidence (see the **Supplementary Note** for estrogen receptor and RNA PolII binding data at both the stringent and a lower threshold). The stringent threshold represents a false discovery rate (FDR) of ~1%, and the lower threshold represents an FDR of ~5%. After BLAT analysis[15] to eliminate redundant sequences, we identified a final set of 3,665 unique estrogen receptor binding sites and 3,629 unique RNA PolII binding sites using the stringent threshold, resulting in an estrogen receptor and RNA PolII binding site on average every 839 kb and 847 kb in the genome, respectively.

Correlation of binding with transcription start sites

We mapped the relative location of estrogen receptor and RNA PolII binding sites to transcription start sites (TSS) of known genes from RefSeq (**Fig. 1a**). Approximately 67% of RNA PolII sites map to promoter-proximal (–800 bp to +200 bp) regions of known genes, consistent with findings reported for transcription factor IID (TFIID)[16]. Identification of essential elements for estrogen receptor–mediated transcription of target genes have focused primarily on promoter-proximal regions, and recent estrogen receptor location analyses analyzed only promoter regions[9,10]. However, in our complete genome-wide approach, we find that only 4% of estrogen receptor binding sites mapped to 1-kb promoter-proximal regions at either the high or low threshold (**Fig. 1a**), and as such, almost all *in vivo* estrogen receptor binding events occur in regions previously unannotated as *cis*-regulatory elements within the genome. The low frequency of promoter-proximal binding sites found for estrogen receptor is unlikely to be due to a bias in the method, as we were able to find the vast majority of RNA PolII binding sites at promoters using this method as expected. However, within the list of estrogen receptor binding sites near promoter-proximal regions, we found a number of previously identified estrogen receptor targets, including *TFF1*, *EBAG9*, *TRIM25* (also known as *Efp*), *ESR1* and prothymosin α (*PTMA*), found using the stringent threshold, and cathepsin D (*CTSD*), *PGR* (also known as *PR*), keratin 19 (*KRT19*), *RARA* (also known as *RARα*) and *HSPB1* (also known as *Hsp27*), found using the more relaxed threshold (reviewed in refs. 17,18). Even when a very relaxed cutoff was analyzed corresponding to an FDR of >50%, only three additional promoter-proximal regions previously suggested to be estrogen receptor targets were identified (**Supplementary Table 1** online). The promoters identified using the lower thresholds may

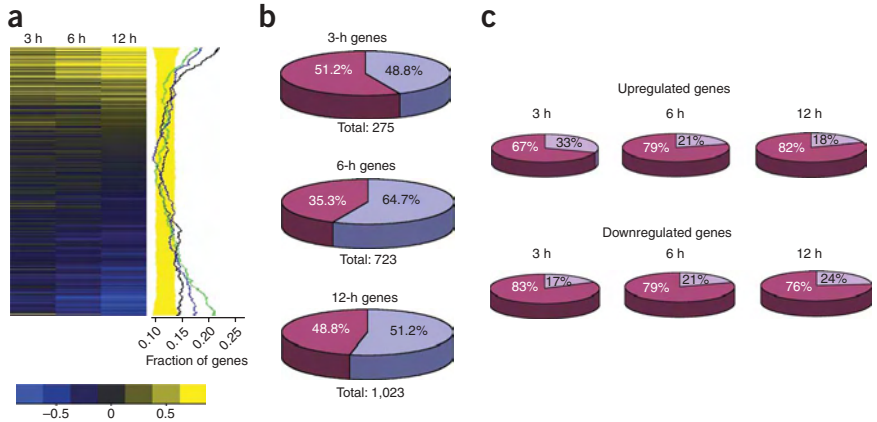


Figure 2 Estrogen-mediated transcript changes and correlation with estrogen receptor binding. (**a**) Expression changes of all genes as ranked by Welch *t* statistic at 3, 6 and 12 h relative to 0 h. Induction of gene expression relative to 0 h is represented as yellow and repression as blue. The graph represents the fraction of genes with an estrogen receptor binding site within 50 kb of the transcription start site. Genes were ranked by Welch *t* statistic between 3, 6 and 12 h and 0 h (control). The black (3 h), blue (6 h) and green (12 h) lines represent 2,000 gene moving averages of the fraction of genes that have one or more estrogen receptor binding sites within 50 kb of the transcription start site. The yellow band is a 99% confidence interval for the binding site moving average of genes in the 25%–50% 12-h *t* statistic range. (**b**) Summary of estrogen-mediated expression changes over a time course (0, 3, 6 and 12 h). Shown are the number of differentially expressed genes after estrogen treatment, relative to the vehicle-treated control (0 h). Blue segments represent upregulated genes, and red segments represent downregulated genes. (**c**) Percentage of genes upregulated or downregulated at each time point (relative to time 0 h) that contain estrogen receptor binding sites within 50 kb (purple sector).

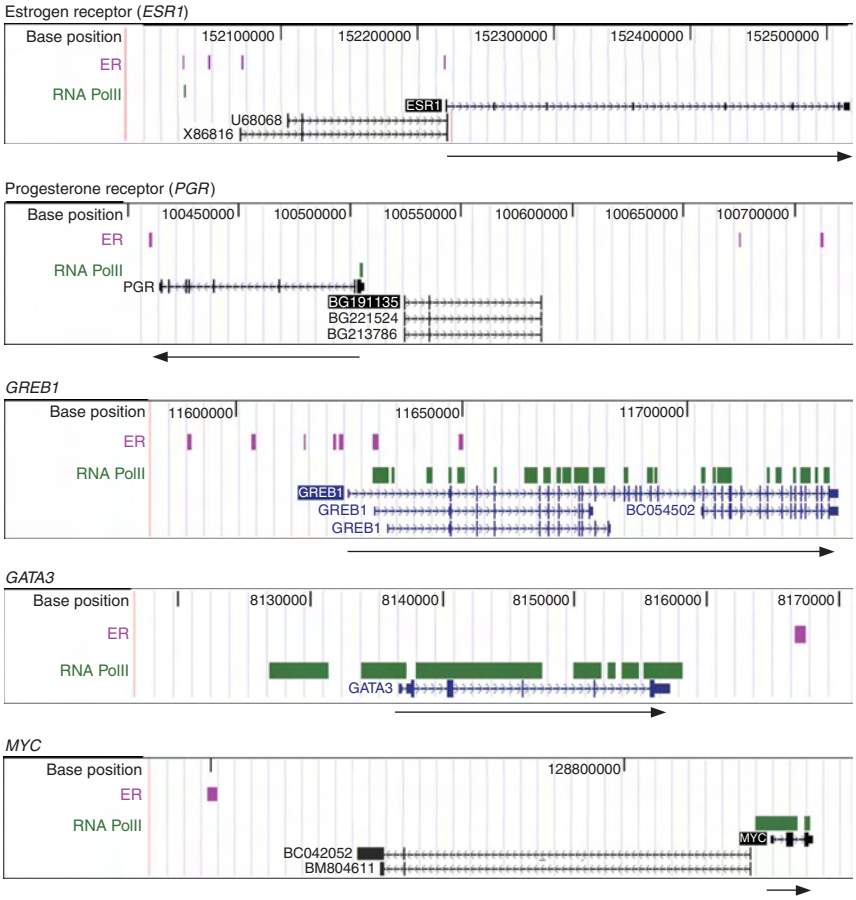


Figure 3 Estrogen receptor and RNA PolII binding relative to specific gene targets. The purple blocks represent estrogen receptor (ER) binding sites, and green blocks represent RNA PolII sites. *ESR1*, *GREB1*, *MYC* and *GATA3* are shown in their genuine 5′-3′ orientation, and *PGR* is shown in its genuine 3′-5′ orientation. The black arrows indicate the direction of the gene. Included are predicted transcripts that exist between the estrogen receptor binding sites and the target genes.

represent indirect or secondary binding sites, as assessed by the low enrichment (1.2- to 1.8-fold over background) by directed quantitative ChIP (**Supplementary Fig. 1** online and data not shown), compared with 15- to 160-fold for adjacent estrogen receptor binding sites distal from promoter regions.

Conserved *cis* elements define estrogen receptor binding

RNA PolII binding correlated well ($r^2 = 0.88$) with gene number, not chromosome length ($r^2 = 0.29$), as its binding was predominately promoter proximal (**Fig. 1b**). Compared with RNA PolII, estrogen receptor binding was less well correlated with gene number ($r^2 = 0.62$) and equally correlated with chromosome size, as estrogen receptor binding is distributed within and between genes rather than being restricted to promoters (**Fig. 1b**).

Sequence comparison of all the estrogen receptor binding sites between the genomes of multiple vertebrate species showed high conservation within the binding sites, but not in immediate surrounding regions (**Fig. 1c**); conservation was almost to the same level as for coding sequences. Conservation analysis of RNA PolII binding sites showed a similar degree of sequence preservation, although in contrast to estrogen receptor, this was also maintained in the surrounding coding sequence (**Fig. 1c**). Therefore, the evolutionary maintenance of the estrogen receptor binding sites supports their putative role as functional *cis*-regulatory domains distinct from promoters.

Gene expression correlates with binding

To correlate estrogen receptor and RNA PolII binding data with the estrogen transcriptional response, we performed gene expression profiling by microarray analyses, which were performed in triplicate over an estrogen stimulation time course (0, 3, 6 and 12 h), with 3 h representing immediate transcriptional targets[7] and both 6 and 12 h representing delayed targets (complete data sets are available; see **Supplementary Note**). Relative to time 0 h, 134 genes were upregulated after 3 h of estrogen treatment (**Fig. 2a**,**b**), which is a small fraction of the RNA PolII binding sites present in MCF-7 cells under these conditions. However, RNA PolII binding sites identified by ChIP-on-chip represent not only the genes differentially regulated by estrogen, but also estrogen-independent binding sites within actively transcribed genes[13].

Correlation of estrogen receptor binding sites with early (3 h) and late (6 h and 12 h) estrogen-induced genes showed a bias of binding sites within 50 kb of TSS of both early and delayed estrogen-induced genes ($P < 0.001$) (**Fig. 2a**,**c**). Although there is significantly greater estrogen receptor binding bias toward early upregulated genes, the bias observed near late-upregulated genes suggests that either these late transcripts are produced early and do not accumulate to detectable levels for more than 3 h, or more likely, their transcription requires estrogen induction of a secondary or cooperating transcription factor.

Estrogen-mediated gene repression

Most work investigating estrogen-regulated transcription focuses on upregulated genes, although downregulated genes constitute a significant fraction of all estrogen-dependent expression changes in cell lines[19] and tumor samples[20]. In our expression array analysis, 51.2% of early (3 h) gene changes are downregulated events (**Fig. 2b**). Of the different possible mechanisms for this early gene inhibition, one hypothesis is a sequestration of limiting factors away from downregulated genes[21], so-called physiologic squelching. In support of this hypothesis, correlation of estrogen receptor binding sites with downregulated genes did not show any statistical bias to the TSS of genes downregulated at 3 h (**Fig. 2a**). We took several different experimental approaches to assess if physiologic squelching was a primary mode of early downregulation. RNA PolII binding at the promoters of early-downregulated genes decreased after only 45 min of estrogen stimulation, coincident with RNA PolII binding at promoters of early-upregulated genes (data not shown). Furthermore, pretreatment of MCF-7 cells with the translational inhibitor cycloheximide for 1 h before estrogen stimulation did not influence the early decreases in a number of assessed transcripts (**Supplementary Fig. 2** online), suggesting that these genes are primary, yet indirect, targets of estrogen receptor action.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

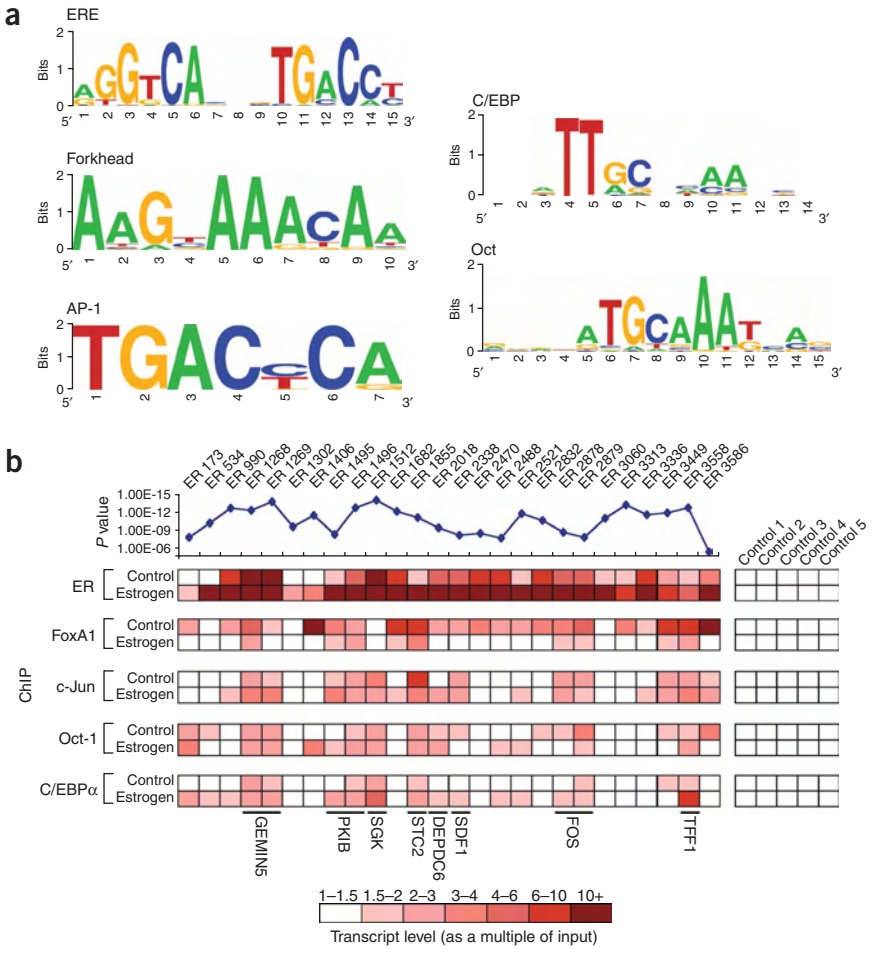


Figure 4 Identification of enriched motifs within the estrogen receptor binding sites and validation of transcription factor binding. (**a**) A computational screen for enriched motifs within all estrogen receptor binding regions demonstrates the presence of ERE, Forkhead, AP-1, Oct and C/EBP sites, with nucleotide bias shown using Weblogo (http://weblogo.berkeley.edu/). A complete list of enriched motifs can be found in **Supplementary Table 2**. (**b**) Directed ChIP of transcription factors that bind to these enriched motifs was performed on 26 estrogen receptor (ER) binding sites and five control regions. The binding sites were chosen to cover a range of enrichment values but also included sites near a select number of estrogen-regulated genes. The relative *P* value for each of the binding sites assessed is provided. Estrogen receptor binding sites adjacent to estrogen-regulated genes are shown by the gene name. The real-time PCR data are shown as a multiple of input DNA and are the average of independent replicates.

In contrast to the early-downregulated genes, when we mapped the relationship between estrogen receptor binding and the TSS of genes downregulated at the later 6- and 12-h time points, we observed a significant enrichment of estrogen receptor binding sites within 50 kb of promoter regions (**Fig. 2a**). This bias of estrogen receptor binding adjacent to late-downregulated genes suggests that in contrast to the majority of early-downregulated genes, which are likely to result from a preponderance of indirect mechanisms such as physiologic squelching, most downregulation late requires estrogen receptor binding. The lag suggests the necessity for the transcription of an estrogen-induced repressor or corepressor capable of associating with chromatin-bound estrogen receptor to facilitate subsequent transcriptional inhibition of adjacent genes. In support of this hypothesis, pretreatment of MCF-7 cells with cycloheximide before estrogen stimulation abrogated the late downregulation of a number of assessed transcripts (**Supplementary Fig. 2**), confirming the requirement for translation of a secondary factor.

Diversity of estrogen receptor regulatory mechanisms

The ChIP-on-chip data suggest that a diversity of binding profiles exist. As examples, autoregulation of the *ESR1* gene involved estrogen receptor binding at the promoter as previously implicated[22] but also may involve three estrogen receptor binding sites 150 kb to 192 kb upstream of the gene (**Fig. 3**). The gene encoding the progesterone receptor, a steroid receptor that is critical in female reproduction and lactation[23] and pathological in breast cancer, contained a RNA PolII binding site at the promoter and two estrogen receptor binding sites 168 kb and 206 kb upstream of the gene. In contrast, approximately half of early, direct estrogen-upregulated genes have estrogen receptor binding sites within 100 kb. As examples, *GREB1*, an estrogen-regulated gene[24] with no previously identified mechanism of estrogen regulation, contained RNA PolII and an estrogen receptor binding site at the promoter of a specific isoform, as well as a cluster of five other estrogen receptor sites upstream of the gene. *GATA3*, a transcription factor that correlates with estrogen receptor status in breast cancer cells[25], contained one estrogen receptor binding site close to the 3′ end of the gene. Previous work delineating mechanisms of estrogen induction of *MYC* have implicated non-estrogen-responsive elements (EREs) within the promoter[26] as the estrogen receptor binding site[7], but we observed a single estrogen receptor binding site approximately 67 kb upstream from *MYC*. We validated estrogen receptor binding to most of this subset of newly identified binding sites using directed estrogen receptor ChIP and real-time PCR (**Supplementary Fig. 1**). In support of the ChIP-on-chip data, estrogen receptor binding was only marginally enriched at the *MYC* promoter by ChIP and quantitative PCR (1.5-fold over input DNA) compared with the newly identified upstream enhancer (15-fold over input DNA), substantiating the assertion that the *MYC* promoter is not the primary estrogen receptor binding site. It should be noted that in the cases of *ESR1*, *PGR* and *MYC*, predicted transcripts exist in the region between the binding sites and the hypothesized target, as shown in **Figure 3**, although there is no evidence for their expression in MCF-7 cells. Future studies will be needed in order to prove the particular functional significance of any of these estrogen receptor binding sites; however, in the absence of this unique resource, the existence of these sites would be unknown. These examples typify the gene-specific complexity of estrogen receptor transcriptional regulation and reinforce the concept that the historical bias towards promoter-proximal regions does not fully identify the primary sites of estrogen regulation in most cases.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

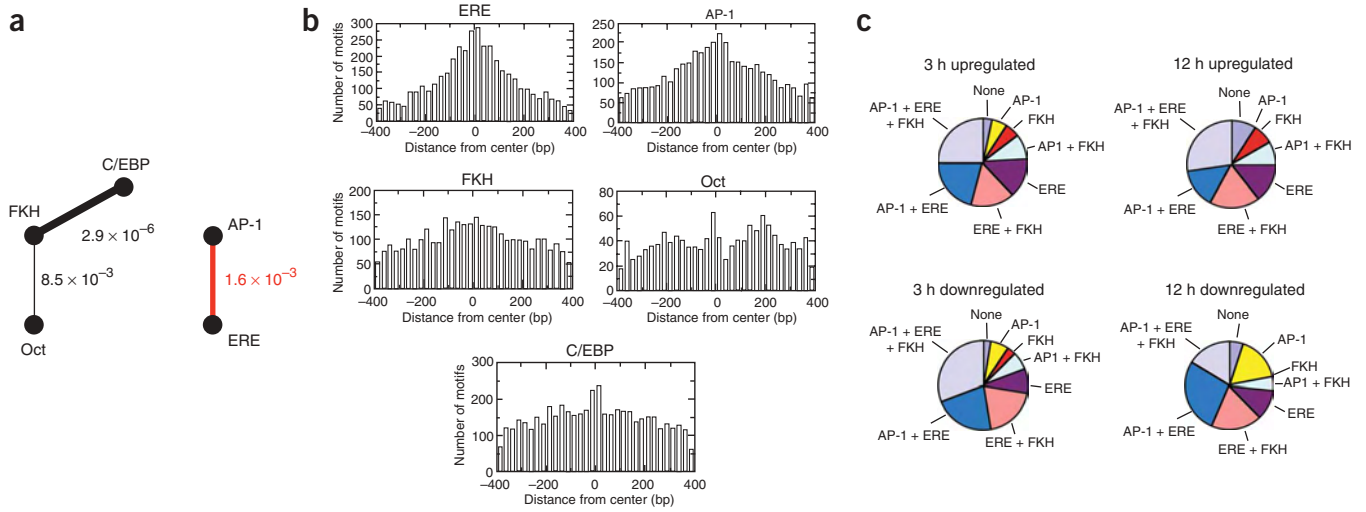


Figure 5 Involvement of cooperating transcription factors at estrogen receptor binding sites. (**a**) Pairwise analysis between ERE, Forkhead (FKH), AP-1, Oct and C/EBP motifs. A positive correlation is shown as a black line, and a negative correlation is shown as a red line. Statistical significance is shown numerically and also indicated by line thickness. (**b**) Distribution of ERE, Forkhead (FKH), AP-1, Oct and C/EBP motifs within estrogen receptor binding sites relative to the center of the binding sites (represented as 0). (**c**) Fraction of specific binding sites containing ERE, AP-1 and Forkhead (FKH) motifs adjacent to genes up- or downregulated early (3 h) or late (12 h). The top 200 differentially expressed genes at each time point (based on the Welch *t* test) were included in the analysis. For each gene, only motifs in the nearest ChIP region within 50 kb were considered.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Involvement of cooperating factors

To systematically identify the network of transcription factors that modulate estrogen receptor function, we searched all estrogen receptor binding sites for enriched DNA binding elements by both *de novo* and candidate scanning approaches. This screen identified EREs and Forkhead motifs, as previously implicated[11], as well as a number of other putative binding motifs (a complete list of enriched motifs can be found in **Supplementary Table 2** online), including AP-1, Oct and C/EBP motifs (**Fig. 4a**), supporting the suggestion that these sites serve as enhancers. Using ChIP followed by real-time PCR of 15 randomly selected estrogen receptor binding sites with different enrichment values, 11 sites adjacent to estrogen regulated genes and five negative controls (regions containing EREs or ERE half sites, but not identified as estrogen receptor binding sites) (**Supplementary Table 3** online), we confirmed estrogen receptor recruitment to all of the tested ChIP-on-chip–identified sites but not to any of the negative controls (**Fig. 4b**). FoxA1 binding occurred at most of these sites (but not at any of the controls), and the signal was generally diminished after estrogen addition, as we previously found for sites on chromosomes 21 and 22 (ref. 11) (**Fig. 4b**).

To validate specific transcription factor association with the enriched AP-1, Oct and C/EBP motifs, we focused initially on members of each transcription factor family that were abundant in MCF-7 cells. As an example, Oct-1 was expressed in MCF-7 cells, and Oct-1 protein was shown by ChIP to be recruited to a number (73%) of the assessed sites (**Fig. 4b**), supporting the data showing Oct-1 as a nuclear receptor–interacting transcription factor[27] and a putative regulator of estrogen target genes[28]. Similarly, c-Jun and C/EBPα were shown to bind to a subset of estrogen receptor binding sites, but not to the negative controls. C/EBPα has been shown to interact with estrogen receptor in GST pull-down experiments[29], and c-Jun has an extensively characterized role modulating estrogen target genes[30,31], although general roles for these transcription factors in estrogen receptor–mediated transcription have not been previously shown. Importantly, these motifs were not statistically enriched in the promoter-proximal regions of estrogen-regulated genes (data not shown).

We performed pairwise analysis to identify combinatorial interactions between ERE, Forkhead, Oct, AP-1 and C/EBP motifs within all estrogen receptor binding sites and found a strong negative correlation between ERE and AP-1 elements (**Fig. 5a**), suggesting that ERE and AP-1 motifs occur exclusively. The pairwise analysis also showed a positive correlation between C/EBP, Oct and Forkhead motifs (**Fig. 5a**), implying that these motifs tend to cluster together within the same estrogen receptor binding sites. The C/EBP, Oct and Forkhead motif cluster had equal likelihood of occurring with ERE or AP-1 motifs.

The relative positional distribution of the enriched motifs within the estrogen receptor binding sites show that both ERE and AP-1 motifs typically occur at the center of the estrogen receptor binding sites (**Fig. 5b**), whereas Forkhead, C/EBP and Oct motifs were less biased toward the center of the binding sites, possessed a more even distribution across the estrogen receptor binding sites and, in the case of Oct motifs, seemed to be multimodal, with clusters occurring approximately 200 bp on both sides of the center of the binding sites. This suggests that the primary interaction of estrogen receptor with chromatin can occur either through direct interaction with an ERE or through a tethering mechanism involving AP-1 factors, as previously suggested[18,31], with C/EBP, Oct and Forkhead[32,33] motifs functioning as adjacent binding sites for cooperating factors.

NRIP1-mediated gene repression

We next investigated whether there were functional differences between estrogen receptor binding sites centered on an ERE versus an AP-1 motif in binding sites adjacent to the highest differentially regulated genes. In contrast to the early-regulated genes, there was a clear bias of AP-1–centered estrogen receptor binding sites adjacent to late (12 h)-downregulated versus late-upregulated genes ($P < 0.01$; **Fig. 5c**). As this bias in AP-1 motifs was not observed early, it suggested that the late direct estrogen receptor binding–mediated

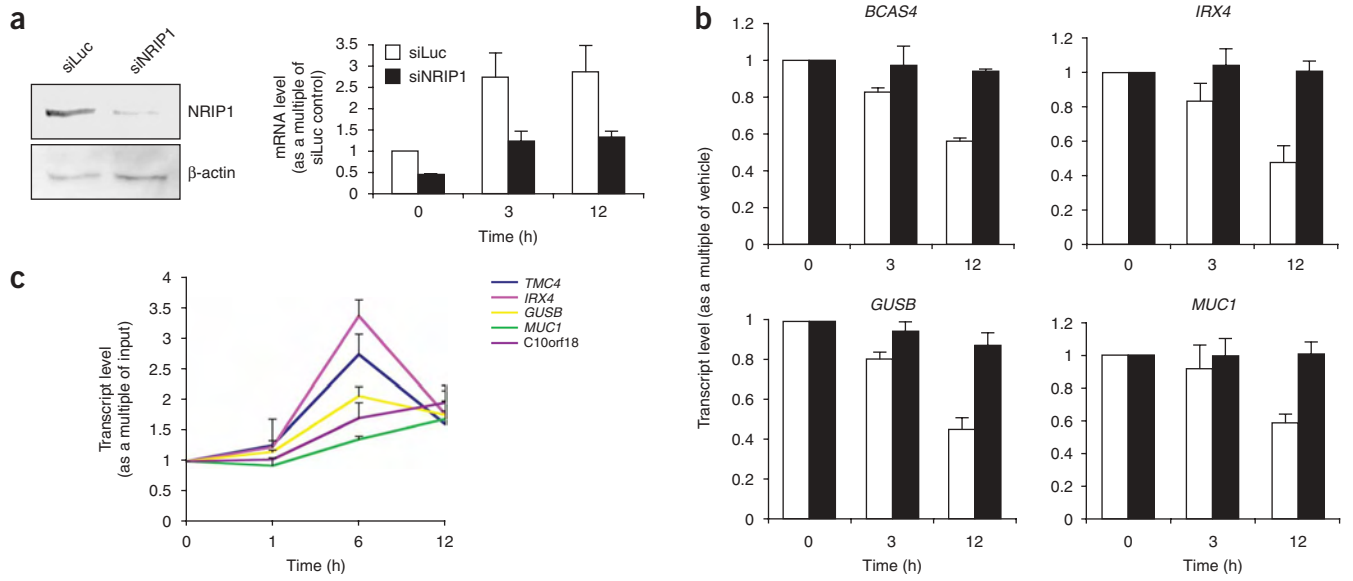


Figure 6 The role of NRIP1 in mediating gene repression. (**a**) siRNA to control (siLuc) or NRIP1 was transfected into hormone-depleted MCF-7 cells, and NRIP1 protein levels were assessed (left). β-actin functioned as a loading control. *NRIP1* mRNA levels were assessed after estrogen stimulation in the presence of control (siLuc) or siNRIP1. (**b**) Transcript levels of candidate late-downregulated genes with estrogen receptor binding sites containing AP-1 elements (*BCAS4*, *IRX4*, *GUSB* and *MUC1*) were assessed after siLuc control or siNRIP1 transfection and subsequent estrogen stimulation. The data are normalized to vehicle-treated conditions. (**c**) We assessed NRIP1 recruitment to estrogen receptor binding sites containing AP-1 elements adjacent to late-downregulated genes by NRIP1 ChIP after estrogen treatment for increasing time periods. Real-time PCR was performed on the estrogen receptor binding sites and data were normalized to vehicle-treated conditions. The data are the mean of independent replicates ± s.d.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

transcriptional inhibition might be mediated via an estrogen-induced factor capable of interaction with estrogen receptor tethered to AP-1 motifs.

We therefore searched for genes that were estrogen induced at the early (3 h) time point that were known to interact with either estrogen receptor or AP-1 proteins. One such candidate was the coregulator NRIP1, which (i) is upregulated at 3 h of estrogen treatment, (ii) is a nuclear receptor corepressor[34] and (iii) has been shown *in vitro* to specifically antagonize estrogen receptor transcription via its interaction with AP-1 proteins[35].

To assess whether NRIP1 was a required factor mediating late, direct gene repression via estrogen receptor binding to AP-1–containing elements, we developed short interfering RNA (siRNA) to the *NRIP1* transcript and transfected this into hormone-depleted MCF-7 cells. NRIP1 protein levels were effectively reduced after specific siRNA transfection, and the early estrogen-induced accumulation of *NRIP1* transcript in control siRNA-treated cells was significantly inhibited by the presence of siNRIP1 (**Fig. 6a**).

We next measured transcript levels by quantitative RT-PCR of several late (12 h after estrogen treatment) downregulated genes that contained adjacent estrogen receptor binding sites centered on AP-1 elements, including *BCAS4*, *IRX4*, *GUSB* and *MUC1*. All of these target genes were substantially downregulated at 12 h by estrogen, but these effects were markedly reversed in the presence of siNRIP1 (**Fig. 6b**), demonstrating that NRIP1 is necessary for the downregulation of these genes. We found that a number of control target genes that are upregulated late by estrogen were unaffected by the presence of siRNA to *NRIP1* (data not shown). Furthermore, NRIP1 ChIP followed by real-time PCR of the estrogen receptor binding sites adjacent to these late-downregulated genes confirmed NRIP1 binding at either 6 or 12 h of estrogen treatment (**Fig. 6c**).

Function of binding sites in human breast cancers

In order to determine whether the estrogen receptor binding sites defined in MCF-7 cells is cell line specific, we assessed the function of a subset of estrogen receptor binding sites in another estrogen receptor–positive breast cancer cell line, T47D. All of the small subset of tested sites functioned as estrogen receptor binding sites in another breast epithelial cell line (**Fig. 7a**).

To test whether the estrogen receptor binding sites as defined in MCF-7 cells are relevant to the pattern of gene expression observed in authentic human breast cancers, we compared the estrogen receptor binding with the gene expression signatures from two independent studies, one involving 286 primary breast tumors[4] and the other 295 breast tumors[3]. When we compare the position of an estrogen receptor binding site with the genes correlated with estrogen receptor expression in each of the two studies we find a significant (Wang, $P < 3.0 \times 10^{-8}$, and van de Vijver, $P < 1.0 \times 10^{-6}$) enrichment of estrogen receptor binding adjacent to the positively correlated genes (**Fig. 7b**). The percentage of genes with estrogen receptor binding sites within 100 kb are 56% and 59% for the van de Vijver and Wang studies, respectively. This relationship is very similar to the one found for estrogen-regulated (3 h) genes in MCF-7 cells of ~50%. As a comparison, we examined estrogen receptor binding profiles adjacent to estrogen-regulated genes in MCF-7 cells (**Fig. 7c**). This result suggests that the estrogen receptor binding profile identified in MCF-7 cells both predicts the gene expression signature and identifies functional regions of the genome that control estrogen responses in primary human breast cancers.

DISCUSSION

The identification of the set of *cis*-acting targets of a *trans*-acting factor such as the estrogen receptor across the whole genome provides an important new resource for the study of gene regulation. The classic

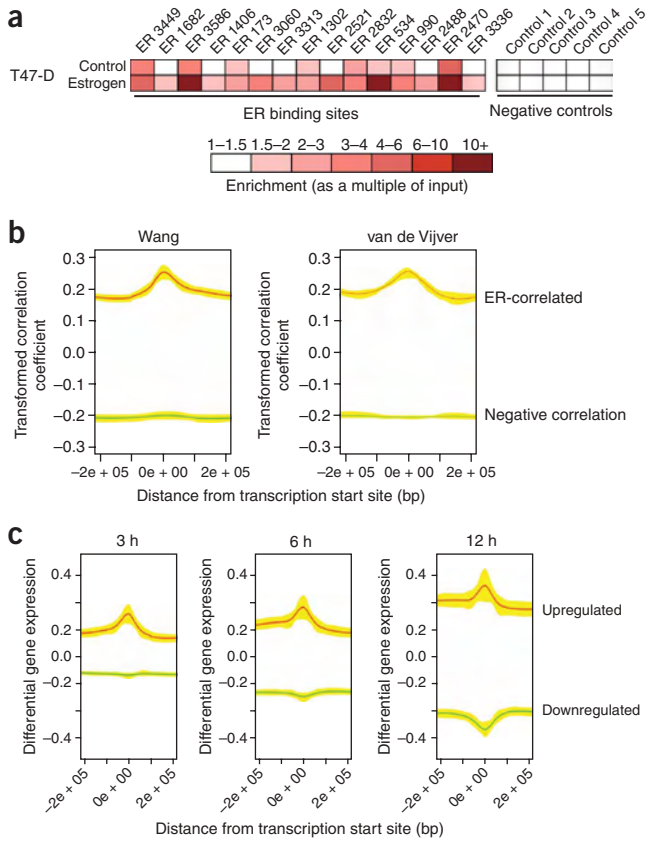


Figure 7 Assessment of estrogen receptor binding properties in different cell systems. (**a**) Estrogen receptor (ER) ChIP was performed after vehicle or estrogen stimulation in T-47D breast cancer cells. Real-time PCR of estrogen receptor binding sites previously identified in MCF-7 cells was performed, and the data are shown as a multiple of input. (**b**) Correlation of estrogen receptor binding sites relative to transcription start sites of the highest estrogen receptor–correlated genes from two independent primary breast cancer gene expression studies. (**c**) Correlation of estrogen receptor binding sites with transcription start sites of genes either estrogen-upregulated or estrogen-downregulated at 3, 6 or 12 h (relative to 0 h) in MCF-7 cells.

paradigm of estrogen receptor function involves binding to promoter-proximal regions and subsequent gene regulation. However, it now seems that the promoter-proximal regions, although important for some genes, do not constitute the majority of estrogen receptor target sites. Instead, it is apparent that a full definition of estrogen receptor binding to *cis*-regulatory regions distinct from promoters is required to fully understand the estrogen response. Similar analyses of c-Myc, p53 and Sp-1 binding to chromosomes 21 and 22 has also shown analogous enhancer binding profiles[36], suggesting that studies that focus on promoter regions[9,10] are insufficient. In contrast, TFIID[16] and RNA PolII ChIP-on-chip analyses (in this investigation) confirm that the basal transcription machinery is significantly biased to promoter-proximal regions. In general, it seems that communication is often mediated at great distances between the transcription factors that initiate gene expression events and the transcription machinery that execute it.

Almost one-third of early–estrogen upregulated genes have estrogen receptor binding sites within 50 kb of the TSS, confirming a clear statistical bias for regulation of genes in the vicinity of chromatin-interaction sites. Other estrogen-stimulated genes that do not have an estrogen receptor binding site within 50 kb may use sites that are greater than 50 kb from the gene[11], use enhancers on different chromosomes[37] or induce transcription independent of estrogen receptor binding events. It is of interest to note that there are many more estrogen receptor binding sites in the genome than differentially regulated genes, as has been previously suggested[38]. It is likely that a significant number of these binding sites are not functional in MCF-7 cells under the specific experimental conditions used and may be functional in other cell types or under different conditions.

Although previous work has shown numerous estrogen receptor–cooperating proteins at the promoters of estrogen-regulated genes[6,8], we find that transcriptional activity of estrogen receptor from the *cis*-regulatory elements also involves combinations of cooperating transcription factors. We previously found an enrichment of Forkhead motifs within estrogen receptor binding sites on chromosomes 21 and 22 and subsequently showed a requirement for FoxA1 in mediating estrogen receptor binding to chromatin[11] supporting the role of FoxA1 as a pioneer factor[32,33]. Using the statistical power of all 3,665 estrogen receptor binding sites in the entire human genome, we both confirmed the role of FoxA1 and identified several additional enriched motifs that were not identified in our previous investigation[11], including DNA-binding motifs for AP-1, C/EBP and Oct transcription factors. Previous work has shown an estrogen-dependent role for c-Jun, Oct-1 and C/EBP proteins in transcription of cyclin D1 (ref. 28), but the unbiased identification of these binding motifs within estrogen receptor binding sites suggests a more general role for these cooperating factors in estrogen receptor transcription.

AP-1 family members have an extensively characterized role in estrogen receptor–regulated transcription[31], and the estrogen receptor can bind to DNA via ERE or AP-1 elements[18,30] involving different protein complexes[39]. A positive role for AP-1 proteins in the estrogen-mediated induction of target genes is established, but we now show a role for AP-1 proteins in gene repression. Our data show that gene changes that occur late (at 6 and 12 h of estrogen stimulation) can be clearly divided into two categories: genes that are upregulated, which have adjacent estrogen receptor binding sites more likely to contain EREs, and genes that are downregulated, which generally contain AP-1 elements. We now show the mechanisms defining these two classes of estrogen receptor binding sites, with estrogen inducing the corepressor NRIP1, which subsequently interacts with estrogen receptor–AP-1 complexes[35] to effect direct repression of adjacent target genes. Our previous work identified the mechanism of estrogen receptor–mediated *NRIP1* induction: several distant enhancers (~150 kb from the TSS of *NRIP1*) function as primary estrogen receptor binding sites, and chromatin loops between these *NRIP1* enhancers and its promoter exist in the presence of estrogen[11].

The estrogen receptor is critical in determining the phenotype of human breast cancers and is the most important therapeutic target. The complete set of estrogen receptor binding sites across the genome defined in these studies establishes a new resource for understanding estrogen action in breast cancer. It correctly predicts the genes coexpressed with the estrogen receptor in primary breast tumors and thus identifies important and previously unexplored regions of the genome that are the critical regulators of the estrogen dependence of breast cancer.

METHODS

ChIP-on-chip analysis. ChIP and chromatin preparation were performed as previously described[11,40,41]. We used antibodies to ERα (Ab-10; Neomarkers,

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Lab Vision); ERα (HC-20) and RNA PolII (H-224) (Santa Cruz) and RNA PolII (4H8; Abcam). All three replicates were performed on the Affymetrix Human tiling 1.0 microarrays (14-chip set). The only difference between replicates is that the Affymetrix imagine software GCOS rotated the CEL files 90° in the first two replicates but not in the third replicate. We applied the generalized Mann-Whitney *U* test[13] to identify regions at least 600 bp in length that were enriched in ChIP samples compared with the controls. A total of 5,712 regions were predicted at the *P* value cutoff of 1×10^{-5}. MAT[14] was applied to the same data to predict the highest-scoring 5,712 ChIP regions (equivalent to a MAT score cutoff of 10.27 and a *P* value of 7.1×10^{-6}). The two predictions had a high degree of concordance, and we reported the intersection between them. In addition, 17 regions predicted by MAT as the top 1,000 but missed by the generalized Mann-Whitney method were added to the final list of estrogen receptor binding sites. BLAT analysis[15] was performed to eliminate redundant sequences.

Expression microarrays. MCF-7 cells were deprived of hormones as previously described[11] and stimulated with 100 nM estrogen for 0, 3, 6 or 12 h, after which total RNA was collected using Trizol (Invitrogen). Expression microarrays were Affymetrix U133Plus2.0, and all experiments were performed in triplicate. Data were analyzed using the RMA algorithm[42] with the newest probe mapping[43], and the Welch *t* statistic was used to calculate the level of differential expression at each time point relative to 0 h.

Directed ChIP and real-time PCR. ChIP was performed as previously described[11]. We used antibodies to ERα (Ab-10; Neomarkers, Lab Vision); ERα (HC-20), HNF-3α/FoxA1 (H-120), c-Jun (N), Oct-1 (C-21), C/EBPα (14AA) and NRIP1 (H-300) (Santa Cruz); and NRIP1 (ab3425; Abcam). Quantitative real-time PCR was performed as previously described[11].

siRNA. siRNA experiments were performed as previously described[11]. *NRIP1* siRNA sequences (Dharmacon) were as follows: sense, 5′-GAAGCGUG CUAACGAUAAA-3′, and antisense, 5′-UUUAUCGUUAGCACGCUUC-3′. Antibodies used in the protein blot were NRIP1/RIP-140 R5027 (Sigma Aldrich) and β-actin A1978 (Sigma Aldrich).

Real-time RT-PCR. RNA was collected as described above. Real-time RT-PCR was performed as described above for real-time PCR, with the exception that 10 units of MultiScribe (Applied Biosystems) were added, and a reverse transcription step of 48 °C for 30 min was included before PCR cycling. Primer sequences can be found in **Supplementary Table 3**.

Sequence conservation analysis. The 3,665 estrogen receptor ChIP regions were aligned at their centers and uniformly expanded to 3,000 bp in each direction, and phastCons[44] scores were retrieved (http://genome.ucsc.edu) and averaged at each position.

Screen of estrogen receptor binding sites for enriched motifs. The ChIP regions and 3,800 promoters of non-differentially expressed RefSeq genes located within 200 kb of the ChIP regions were scanned for transcription factor motifs using 533 well-defined position-specific score matrices (PSSM) from TRANSFAC[45], JASPAR[46] and ref. 11. The background nucleotide frequencies were computed from the whole genome. For each matrix, we considered all PSSM matches with cutoff scores from 5.0 (90% of relative entropy) up to 12.0, in increments of 0.5. At each cutoff level, the resulting two sets of motifs were then tested for significance using the criteria of binomial $P < 1 \times 10^{-4}$ and minimum change (with respect to control) of 1.5-fold. We report the relevant statistics for only those PSSM score cutoffs with maximum changes with respect to control. In addition to the PSSM scan, we performed *de novo* motif scans using LeitMotif [47], a modified MDscan[48] with ninth-order Markov dependency for the genome background.

Conditional independence graphical models[49] were constructed to understand the association of transcription factors. The 3,665 estrogen receptor ChIP regions were uniformly resized to 400 bp in each direction from their centers. PSSM scans for ERE, Forkhead, AP-1 and Oct were performed with 90% of relative entropy (RE) cutoff and for C/EBP at a cutoff of 5.0 because of its very low RE. The PSSM scores were then normalized as (score – RE)/motif length, and when two motifs overlapped, only the motif with higher normalized score was kept. The resulting five-dimensional motif hit contingency table for the distribution of the motifs in estrogen receptor ChIP regions was then analyzed with MIM (http://www.hypergraph.dk) graphical modeling software. Using 100% relative entropy adds one more interaction edge between Oct-1 and C/EBP; the corresponding model is shown in **Supplementary Figure 3** online.

Correlation of estrogen receptor binding to gene expression profiles in tumor samples. We downloaded the gene expression index from 286 lymph node–negative individuals who had not received adjuvant systemic treatment[4] and 295 individuals with either lymph node–negative or lymph node–positive disease[3] from GEO (accession 2034) and http://www.rii.com/publications/2002/nejm.html, respectively. Pearson correlation coefficients of estrogen receptor expression relative to every other UCSC known gene were calculated within the Wang and van de Vijver data sets, respectively. Fisher's transformation of the correlation coefficient, $z = 0.5 \log((1 + c) / (1 - c))$, was fitted to the oriented distance to the nearest estrogen receptor ChIP region. A cubic spline with 11 knots between –1 Mb and +1 Mb with equal numbers of data points between knots was applied to smooth the graph (**Fig. 7b**).

URLs. Data to accompany the **Supplementary Note** can be downloaded from http://research.dfci.harvard.edu/brownlab/datasets/index.php?dir=ER_whole_human_genome/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors would like to thank D. Lynch, M. Lupien, M. Laza and O. Jänne for helpful discussions and review of the manuscript. This work was supported by grants from the US National Institute of Diabetes and Digestive and Kidney Diseases (DK074967) and the US National Cancer Institute–sponsored Dana-Farber/Harvard Cancer Center Breast Cancer SPORE Grant (CA089393); by the Claudia Adams Barr Program in Innovative Basic Cancer Research (M.B., P.A.S., X.S.L.); by US Department of Defense Breast Cancer Research Program Awards (M.B., J.S.C.) and by fellowships from the Fondation Recherche Medicale (J.E.), the Susan G. Komen Breast Cancer Foundation (E.K.K.) and the US National Institutes of Health (T.R.G. and C.A.M.).

AUTHOR CONTRIBUTIONS

This study was designed by J.S.C., C.A.M., A.S.B., P.A.S., X.S.L., T.R.G. and M.B. The experimental procedures were performed by J.S.C., with assistance in specific areas from T.R.G., J.E., E.K.K., K.C.F. and Q.W. Biostatistical support was provided by C.A.M., J.S., W.L. and X.S.L. Microarray support and analysis was provided by G.F.H., S.B., V.S., E.A.F. and T.R.G.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874 (2001).
2. van 't Veer, L.J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
3. van de Vijver, M.J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
4. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).
5. Perou, C.M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
6. Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852 (2000).
7. Shang, Y. & Brown, M. Molecular determinants for the tissue specificity of SERMs. *Science* **295**, 2465–2468 (2002).
8. Metivier, R. *et al.* Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763 (2003).
9. Laganiere, J. *et al.* Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc. Natl. Acad. Sci. USA* **102**, 11651–11656 (2005).
10. Cheng, A.S. *et al.* Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters. *Mol. Cell* **21**, 393–404 (2006).

11. Carroll, J.S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long -range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
12. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154 (2005).
13. Brodsky, A.S. *et al.* Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.* **6**, R64 (2005).
14. Johnson, W.E. *et al.* MAT: Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci. USA* (in the press) (2006).
15. Kent, W.J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
16. Kim, T.H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
17. Klinge, C.M. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res.* **29**, 2905–2919 (2001).
18. O'Lone, R., Frith, M.C., Karlsson, E.K. & Hansen, U. Genomic targets of nuclear estrogen receptors. *Mol. Endocrinol.* **18**, 1859–1875 (2004).
19. Frasor, J. *et al.* Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* **144**, 4562–4574 (2003).
20. Nishidate, T. *et al.* Genome-wide gene-expression profiles of breast-cancer cells purified with laser microbeam microdissection: identification of genes associated with progression and metastasis. *Int. J. Oncol.* **25**, 797–819 (2004).
21. Kalaitzidis, D. & Gilmore, T.D. Transcription factor cross-talk: the estrogen receptor and NF-kappaB. *Trends Endocrinol. Metab.* **16**, 46–52 (2005).
22. Lazennec, G., Huignard, H., Valotaire, Y. & Kern, L. Characterization of the transcription start point of the trout estrogen receptor-encoding gene: evidence for alternative splicing in the 5′ untranslated region. *Gene* **166**, 243–247 (1995).
23. Lydon, J.P. *et al.* Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes Dev.* **9**, 2266–2278 (1995).
24. Ghosh, M.G., Thompson, D.A. & Weigel, R.J. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer Res.* **60**, 6367–6375 (2000).
25. Hoch, R.V., Thompson, D.A., Baker, R.J. & Weigel, R.J. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int. J. Cancer* **84**, 122–128 (1999).
26. Dubik, D. & Shiu, R.P. Mechanism of estrogen activation of c-myc oncogene expression. *Oncogene* **7**, 1587–1594 (1992).
27. Prefontaine, G.G. *et al.* Selective binding of steroid hormone receptors to octamer transcription factors determines transcriptional synergism at the mouse mammary tumor virus promoter. *J. Biol. Chem.* **274**, 26713–26719 (1999).
28. Cicatiello, L. *et al.* Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter. *Mol. Cell. Biol.* **24**, 7260–7274 (2004).
29. Boruk, M., Savory, J.G. & Hache, R.J. AF-2-dependent potentiation of CCAAT enhancer binding protein beta-mediated transcriptional activation by glucocorticoid receptor. *Mol. Endocrinol.* **12**, 1749–1763 (1998).
30. Umayahara, Y. *et al.* Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. *J. Biol. Chem.* **269**, 16433–16442 (1994).
31. Kushner, P.J. *et al.* Estrogen receptor pathways to AP-1. *J. Steroid Biochem. Mol. Biol.* **74**, 311–317 (2000).
32. Cirillo, L.A. *et al.* Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* **17**, 244–254 (1998).
33. Cirillo, L.A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
34. Lee, C.H., Chinpaisal, C. & Wei, L.N. Cloning and characterization of mouse RIP140, a corepressor for nuclear orphan receptor TR2. *Mol. Cell. Biol.* **18**, 6745–6755 (1998).
35. Teyssier, C., Belguise, K., Galtier, F., Cavailles, V. & Chalbos, D. Receptor-interacting protein 140 binds c-Jun and inhibits estradiol-induced activator protein-1 activity by reversing glucocorticoid receptor-interacting protein 1 effect. *Mol. Endocrinol.* **17**, 287–299 (2003).
36. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
37. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. & Flavell, R.A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
38. van Steensel, B. Mapping of genetic and epigenetic regulatory networks using microarrays. *Nat. Genet.* **37**(Suppl.), S18–S24 (2005).
39. Webb, P. *et al.* The estrogen receptor enhances AP-1 activity by two distinct mechanisms with different requirements for receptor transactivation functions. *Mol. Endocrinol.* **13**, 1672–1685 (1999).
40. Ren, B. *et al.* E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev.* **16**, 245–256 (2002).
41. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919 (2002).
42. Irizarry, R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
43. Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33**, e175 (2005).
44. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
45. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
46. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
47. Song, J.S. & Liu, X.S. LeitMotif: *de novo* motif discovery for ChIP-Chip in mammalian genomes. In Preparation.
48. Liu, X.S., Brutlag, D.L. & Liu, J.S. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat. Biotechnol.* **20**, 835–839 (2002).
49. Darroch, J., Lauritzen, S. & Speed, T. Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8**, 522–539 (1980).

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics